

Residual-Bridge Constructs for Conditioned Diffusions

Sean Malory* and Chris Sherlock

*Department of Mathematics and Statistics
Lancaster University
Lancaster LA1 4YF, UK*

Abstract

We introduce a new residual-bridge proposal for approximately simulating conditioned diffusions. This proposal is formed by applying the modified diffusion bridge approximation of Durham and Gallant (2002) to the difference between the true diffusion and a second, approximate diffusion driven by the same Brownian motion, and can be viewed as a natural extension to recent work on residual-bridge constructs (Whitaker et al., 2016). This new proposal attempts to account for volatilities which are not constant and can therefore lead to gains in efficiency over the recently proposed residual-bridge constructs in situations where the volatility varies considerably, as is often the case for larger inter-observation times and for time-inhomogeneous volatilities. These potential gains in efficiencies are illustrated via a simulation study.

1 The Introduction

Diffusions are a flexible class of continuous-time Markov processes whose dynamics are completely characterized by specifying an instantaneous change in mean (henceforth the drift) and an instantaneous variance (henceforth the volatility). This makes them a useful class of processes for building rich models and, as such, they are utilised in many scientific disciplines, including, but not limited to, biology (e.g. Golightly and Wilkinson, 2011), finance (e.g. Ait-Sahalia and Kimmel, 2007), and engineering (e.g. Coffey et al., 2004). In biological applications, and, more generally, in applications involving reaction networks, diffusions are often used as approximate models for the evolution of the numbers of a set of species within a reaction network. In particular, the chemical Langevin diffusion is often used to approximate the chemical master equation (see, for instance, Ethier and Kurtz, 1986; van Kampen, 1992; Wilkinson, 2011; Fearnhead et al., 2014).

A d -dimensional diffusion, X_t , can be defined as the solution to a stochastic differential

*Email: s.malory@lancaster.ac.uk

equation (SDE)

$$dX_t = \mu(X_t, t, \Theta) dt + \sigma(X_t, t, \Theta) dB_t, \quad X_0 = x_0, \quad (1.1)$$

where $t \in [0, T]$, B_t is an r -dimensional standard Brownian motion, and $\Theta \in \mathbb{R}^m$ is a vector of unknown parameters. The drift $\mu : \mathbb{R}^d \times [0, T] \times \mathbb{R}^m \rightarrow \mathbb{R}^d$ corresponds to the infinitesimal change in mean, and the volatility $\zeta := \sigma\sigma^T : \mathbb{R}^d \times [0, T] \times \mathbb{R}^m \rightarrow \mathbb{R}^{d \times d}$ corresponds to the infinitesimal variance in the sense that

$$\mathbb{E}(X_{t+\Delta t} | X_t = x, \Theta = \theta) = x + \Delta t \mu(x, t, \theta) + o(\Delta t), \quad (1.2)$$

$$\text{Var}(X_{t+\Delta t} | X_t = x, \Theta = \theta) = \Delta t \zeta(x, t, \theta) + o(\Delta t), \quad (1.3)$$

where we write $f(\Delta t) = g(\Delta t) + o(\Delta t)$ if and only if

$$\lim_{\Delta t \downarrow 0} (f(\Delta t) - g(\Delta t)) / \Delta t = 0.$$

Both the drift and volatility depend on a vector of unknown parameters, Θ , which has a prior density of $p_0(\theta)$. These parameters (which drive the evolution of X_t) often relate to quantities of interest, such as the birth rate of a species, and in light of sparse, noisy, and partial observations of the diffusion, inference for these parameters, along with paths of the diffusion, can theoretically proceed in a Bayesian framework via the particle MCMC methodology of Andrieu et al. (2010). Such schemes rely on the construction of an unbiased approximation to the likelihood of the observations, π , which is typically obtained through an importance-sampling and, more generally, particle-filtering approach.

Sample paths of the diffusion are infinite-dimensional and therefore, in practice, it is necessary to restrict attention to the construction of finite-dimensional skeleton paths of the diffusion. Moreover, the transition density of a large class of diffusions is intractable and exact simulation (e.g. Beskos et al., 2006) of a skeleton path is impossible for most multivariate diffusions. Therefore, for many diffusions, it is necessary to approximate the transition density along a fine grid of skeletal points by a Gaussian density using an Euler-Maruyama (EM) step.

The efficiency of any particle MCMC scheme depends on the variability of the importance weights. Hence, the construction of proposal densities which are consistent with respect to both the observations and the true diffusion is key to designing computationally efficient algorithms. The forward simulation (FS) proposal of Pedersen (1995) uses the EM approximation to simulate skeleton paths between consecutive observations. Such a proposal can suffer from poor performance, particularly for informative observations, since it simulates paths independently of the observations. The modified diffusion bridge (MDB) of Durham and Gallant (2002) overcomes this deficiency by using an EM approximation to the transition density between the current point of the skeleton and any subsequent point, thus leading to a tractable, Gaussian transition density between consecutive points of the skeleton given the next observation. However, such a proposal performs poorly if sample paths of the diffusion exhibit non-linear dynamics as is often the case over relatively large inter-observation times. Lindström (2012) tackles this issue

by constructing a proposal which is a mixture between the FS approach and the MDB approach. The downsides of such a proposal are that, firstly, it needs careful tuning, and, secondly, it is not clear how the proposal behaves as the mesh of the partition tends towards zero. These drawbacks also hold for the proposal of Fearnhead (2008) which comprises of a mixture between the FS approach and an approach which simulates from the stationary distribution of the diffusion (when it exists). Schauer et al. (2013) take a different approach and consider the form of the SDE satisfied by the diffusion conditioned on the next observation; this, in general, has the same volatility as the unconditioned diffusion and an extra term in the drift (Rogers and Williams, 2000, chapter IV, section 39) which *guides* the diffusion towards the observation. This extra term depends on the transition density of the unconditioned diffusion and thus, typically, needs to be approximated by the transition density of a tractable diffusion before forward simulation of a skeleton path (via the EM approximation) can proceed. Unfortunately, implementing such an approach in a statistically efficient way can lead to a computationally expensive algorithm (Whitaker et al., 2016).

The novel proposal introduced in this paper can be seen as a natural extension to the residual-bridge constructs of Whitaker et al. (2016) who propose improving on the MDB approach by: constructing a deterministic path which captures the non-linear dynamics of the diffusion, applying the MDB approximation to the residual process defined as the difference between the true diffusion and this path, and then adding the path back on. An appropriate choice of the deterministic path results in a residual whose dynamics are more linear and thus a proposal density which is closer to the true transition density. It is shown empirically in Whitaker et al. (2016) that, for several diffusions, this proposal, when implemented within a Metropolis-Hastings (MH) importance sampler leads to a larger empirical acceptance probability than a MH importance sampler which uses either the MDB or the construct introduced by Lindström (2012) as a proposal distribution. Furthermore, this empirical acceptance probability is similar to the empirical acceptance probability of a MH importance sampler which uses the guided proposals of Schauer et al. (2013) as a proposal distribution but is achieved with a considerably smaller computational cost. However, this residual-bridge approach, while accounting for the variability in the drift, does not account for the variability in the volatility and can, therefore, perform poorly in scenarios where the volatility varies substantially. This is often the case for larger inter-observation intervals, where the diffusion itself moves substantially over the state space, and for diffusions whose volatility is time-inhomogeneous. The proposal introduced in this paper generalizes the residual-bridge proposals of Whitaker et al. (2016) by applying the approximation used in the MDB to the difference between the true diffusion and a second, carefully chosen, approximate diffusion which is coupled with the original diffusion via the same driving Brownian motion. By attempting to account for the variability in the volatility, this new proposal can lead to greater statistical efficiency in situations where the volatility varies considerably.

2 Conducting Inference for Diffusions

Let X_t be a d -dimensional diffusion satisfying (1.1). Consider the pre-defined sequence of times,

$$\{(t_0, \dots, t_I) \in [0, T]^{I+1} : 0 =: t_0 < t_1 < \dots < t_I := T\}.$$

We have noisy observations, $(y_{t_1}, \dots, y_{t_I}) \in \mathbb{R}^{r \times I}$, of the diffusion at times (t_1, \dots, t_I) such that, for any $i \in \{1, \dots, I\}$,

$$(Y_{t_i} | X_{t_i} = x) \sim N(P_i x, \Sigma_i),$$

where $P_i \in \mathbb{R}^{r \times d}$, and $\Sigma_i \in \mathbb{R}^{r \times r}$ is symmetric and positive semi-definite. Denote the density of the i -th observation by $g_i(y_{t_i} | x_{t_i})$ and between any two consecutive times, t_i and t_{i+1} , define an equispaced partition, $\mathcal{P}_{\Delta t}^i$, to be the set

$$\{(t_{i[0]}, \dots, t_{i[K_i]}) \in [t_i, t_{i+1}]^{K_i+1} : t_i =: t_{i[0]} < \dots < t_{i[K_i]} := t_{i+1}\}$$

such that, for all $j \in \{0, \dots, K_i\}$, $t_{i[j]} := t_i + j\Delta t$ with $\Delta t > 0$ and small. For convenience denote any variable $\psi_{t_{i[j]}}$ by ψ_j^i with $\psi_{t_{i[0]}}$ denoted by ψ^i so that, for instance, $y_{t_{i[0]}} = y^1$ is the first observation, and $x_{K_I}^I = x_T$ is the value of the path at the final time point. Denote the transition density of the diffusion by

$$f_\theta^{s,t}(x|z) := \lim_{\epsilon \downarrow 0} \mathbb{P}(X_t \in [x, x + \epsilon) | X_s = z, \Theta = \theta) / \epsilon^d,$$

where

$$[x, x + \epsilon) := \{v \in \mathbb{R}^d : x_i \leq v_i < x_i + \epsilon \text{ for all } i \in \{1, \dots, d\}\}.$$

Interest lies in $\pi(\theta, x_{\mathcal{P}_{\Delta t}} | y^{1:I})$ which is the posterior density for Θ and the skeleton path defined at the points of $\mathcal{P}_{\Delta t} := \mathcal{P}_{\Delta t}^0 \cup \dots \cup \mathcal{P}_{\Delta t}^{I-1}$ and is proportional to

$$\underbrace{\pi_0^\theta(\theta)}_{\text{Prior for } \theta} \underbrace{\pi_0^{x^0}(x^0)}_{\text{Prior for } x^0} \prod_{i=0}^{I-1} \left(\underbrace{g_{i+1}(y^{i+1} | x^{i+1})}_{\text{Observation density}} \overbrace{\prod_{k=1}^{K_i} f_\theta^{t_{i[k-1]}, t_{i[k]}}(x_k^i | x_{k-1}^i)}^{\text{Density of path between obs.}} \right).$$

The transition density for most diffusions is intractable and exact simulation techniques (e.g. Beskos et al., 2006) are primarily limited to diffusions which, under a suitable transformation, have unit volatility and, therefore, are typically only applicable to one-dimensional diffusions. Hence, for small $\Delta t > 0$, it is usual to make the following Euler-Maruyama (EM) approximation; $f_\theta^{t, t+\Delta t}(x|z) \approx \hat{f}_\theta^{t, t+\Delta t}(x|z)$, where we define

$$\hat{f}_\theta^{t, t+\Delta t}(x|z) := \phi(x; z + \Delta t \mu(z, t, \theta), \Delta t \zeta(z, t, \theta)),$$

with $\phi(x; m, \Psi)$ denoting the density of a Gaussian random variable whose mean and variance are m and Ψ respectively. We consider the corresponding approximate posterior, $\hat{\pi}$, which is proportional to

$$\pi_0^\theta(\theta) \pi_0^{x^0}(x^0) \prod_{i=0}^{I-1} g_{i+1}(y^{i+1} | x^{i+1}) \prod_{k=1}^{K_i} \hat{f}_\theta^{t_{i[k-1]}, t_{i[k]}}(x_k^i | x_{k-1}^i).$$

This approximation introduces a bias which decreases as Δt decreases. Therefore a good proposal must be consistent with the diffusion for any small $\Delta t > 0$. Provided care is taken to construct a scheme which does not mix poorly, using, for example, ideas in Golightly and Wilkinson (2008), inference for this approximate target can proceed via the particle marginal Metropolis-Hastings methodology of Andrieu et al. (2010). Such a scheme involves iterating over different values of θ and through the observations y^1, \dots, y^I . To simplify notation we henceforth drop θ , and to simplify exposition, and the subsequent simulation study, we fix x^0 and consider only one observation at time T . We emphasise that, from a statistical efficiency point of view, *nothing* is lost in making these simplifications since none of the proposals to be discussed in section 2.1 depend on more than the subsequent observation, hence any difference in statistical efficiency for one observation will translate into a similar or greater (due to sequential effects) difference in statistical efficiency over many observations. However, we also emphasise that, by fixing x^0 , we decrease the computational cost of some of the proposals, thereby increasing the apparent computational efficiency of those proposals. With these simplifications the approximate target is

$$\hat{\pi}(x_{1:K}|y^1) \propto g_1(y^1|x_K) \prod_{k=1}^K \hat{f}^{t_{k-1}, t_k}(x_k|x_{k-1}) ,$$

where, for ease of exposition, we have denoted any variable ψ_j^0 by ψ_j , K_0 by K , and t_{0j} by t_j .

2.1 Proposals Based on Diffusion Bridges

For inexact observations, which are the focus of this paper, the particle MCMC methodology requires the sampling of N skeleton paths, denoted by $\{x_{1:K}^{(j)}\}_{j=1}^N$, from a proposal q which is close to $\hat{\pi}$ and the calculation of the normalised importance weights of the form

$$\tilde{w}_j \propto \hat{\pi}(x_{1:K}^{(j)}|y^1)/q(x_{1:K}^{(j)}) . \quad (2.1)$$

The optimal proposal $q^{\text{OPT}} \propto \hat{\pi}$ results in equal weights, however, for most diffusions such a proposal cannot be implemented thus necessitating the need to construct proposals which aim to mimic the optimal proposal. The forward simulation (FS) approach of Pedersen (1995) uses the proposal

$$q^{\text{FS}}(x_{1:K}) \propto \prod_{k=1}^K \hat{f}^{t_{k-1}, t_k}(x_k|x_{k-1}) ,$$

which leads to weights of the form $\tilde{w}_j \propto g(y^1|x_K^{(j)})$. Such a proposal produces paths which are consistent with the true diffusion but which can be inconsistent with the observation since x_K is simulated irrespective of the value of y^1 . Therefore, if the noise in the observation is small the variability of the weights is likely to be large as only a few of the

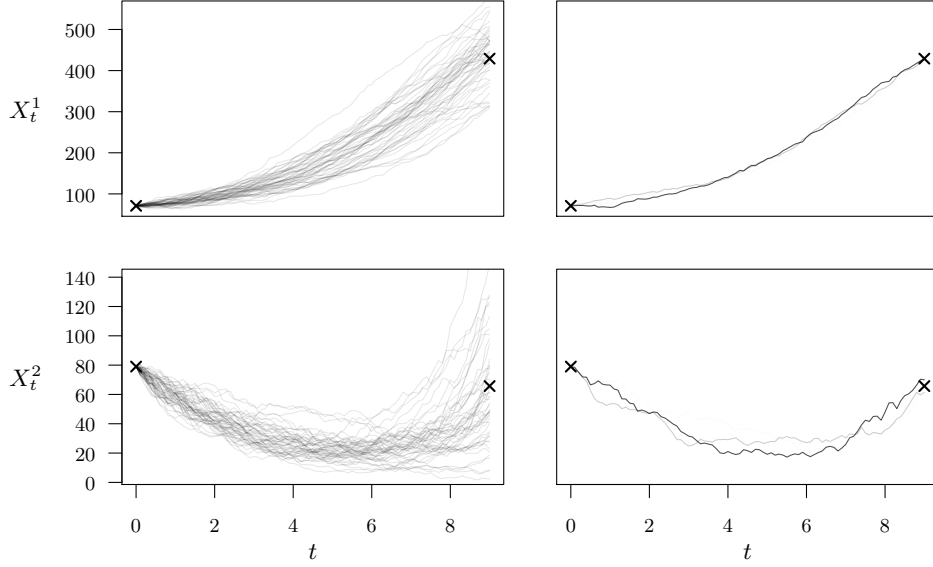


Figure 1: Two sets of plots of fifty paths simulated using the FS approach of Pedersen (1995) on the Lotka-Volterra SDE introduced in subsection 4.1. The plots on the left are the fifty two-dimensional simulated paths with no transparency and the plots on the right are the fifty two-dimensional paths with transparency inversely proportional to their normalised weights so that the path with the largest normalised weight has no transparency and paths with smaller normalised weights are more transparent. The two-dimensional initial condition and observation are illustrated with crosses. It is clear that two of the paths have the highest weight with the other paths having almost zero weight. Moreover, those two paths are precisely the paths whose endpoints lie closest to the observation.

simulated endpoints, $x_K^{(j)}$ will lie near the observation. This phenomena can be seen in figure 1 where we have simulated fifty paths from the Lotka-Volterra SDE introduced in subsection 4.1 using the FS approach of Pedersen (1995). For illustration purposes we have weighted each path under the assumption that the noise in the observation is small¹ and have plotted the paths twice; the paths on the left have no transparency, whereas the paths on the right have been plotted with a transparency inversely proportional to their normalised weights, \tilde{w}_j , so that the path with the largest normalised weight has no transparency and the paths with smaller normalised weights are more transparent. Thus, if there is large variability in the weights the number of partially visible paths will be small, whereas if there is small variability in the weights the number of partially visible paths will be large.

The modified diffusion bridge (MDB) of Durham and Gallant (2002) overcomes the drawback of the FS approach by forming a proposal which depends on the observation y^1 . Specifically, suppose that at time t_k we have simulated x_k . Conditional on this point, form the approximate diffusion, X_t^{MDB} , which satisfies, for $t \in [t_k, T]$,

$$dX_t^{\text{MDB}} = \mu(x_k, t_k) dt + \sigma(x_k, t_k) dB_t, \quad X_k^{\text{MDB}} = x_k. \quad (2.2)$$

¹In particular, for all the figures in this section, we have assumed that $(Y^1|X_K = x) \sim N(x, 5I)$.

This approximation is equivalent to assuming that the EM approximation between the current time point and any subsequent time point is exact and leads to the following joint distribution for the approximate process, X_t^{MDB} , at the next point of the partition and at the observation time;

$$\begin{bmatrix} X_{k+1}^{\text{MDB}} \\ X_K^{\text{MDB}} \end{bmatrix} \Big| (X_k^{\text{MDB}} = x_k) \sim \text{N}(m_k^{\text{MDB}}, \Psi_k^{\text{MDB}}),$$

where

$$m_k^{\text{MDB}} := \begin{bmatrix} x_k + \Delta t \mu(x_k, t_k) \\ x_k + (T - t_k) \mu(x_k, t_k) \end{bmatrix}, \quad \Psi_k^{\text{MDB}} := \begin{bmatrix} \Delta t \zeta(x_k, t_k) & \Delta t \zeta(x_k, t_k) \\ \Delta t \zeta(x_k, t_k) & (T - t_k) \zeta(x_k, t_k) \end{bmatrix}.$$

Consequently, the joint distribution for the approximate process at the next point of the partition and the observation, Y^1 , is given by

$$\begin{bmatrix} X_{k+1}^{\text{MDB}} \\ Y^1 \end{bmatrix} \Big| (X_k^{\text{MDB}} = x_k) \sim \text{N}(\bar{m}_k^{\text{MDB}}, \bar{\Psi}_k^{\text{MDB}}),$$

where

$$\bar{m}_k^{\text{MDB}} := \begin{bmatrix} x_k + \Delta t \mu(x_k, t_k) \\ P_1 x_k + (T - t_k) P_1 \mu(x_k, t_k) \end{bmatrix}, \quad \bar{\Psi}_k^{\text{MDB}} := \begin{bmatrix} \Delta t \zeta(x_k, t_k) & \Delta t \zeta(x_k, t_k) P_1^* \\ \Delta t P_1 \zeta(x_k, t_k) & (T - t_k) P_1 \zeta(x_k, t_k) P_1^* + \Sigma_1 \end{bmatrix},$$

and, in order to avoid confusion with the inter-observation time, T , we have denoted the transpose of a matrix A by A^* . Standard manipulations for the multivariate normal distribution show that

$$(X_{k+1}^{\text{MDB}} | X_k^{\text{MDB}} = x_k, Y^1 = y^1) \sim \text{N}(a_k^{\text{MDB}}, V_k^{\text{MDB}}), \quad (2.3)$$

where

$$\begin{aligned} a_k^{\text{MDB}} &:= x_k + \Delta t \mu(x_k, t_k) + \Delta t \zeta(x_k, t_k) P_1^* ((T - t_k) P_1 \zeta(x_k, t_k) P_1^* + \Sigma_1)^{-1} \\ &\quad \times (y_1 - P_1 x_k - (T - t_k) P_1 \mu(x_k, t_k)), \\ V_k^{\text{MDB}} &:= \Delta t \zeta(x_k, t_k) - \Delta t^2 \zeta(x_k, t_k) P_1^* ((T - t_k) P_1 \zeta(x_k, t_k) P_1^* + \Sigma_1)^{-1} P_1 \zeta(x_k, t_k). \end{aligned}$$

Recall that the approximate process, (2.2), is equivalent to assuming an EM approximation between the current time point and any subsequent time point, hence paths simulated using the MDB exhibit linear dynamics. Thus, even though paths simulated in this way are consistent with the observation, they are inconsistent with any non-linear dynamics of the true diffusion and, hence, can perform poorly in scenarios where the true diffusion exhibits non-linear dynamics and particularly, therefore, for relatively larger values of T . This behaviour, when compared to figure 1, can be seen in figure 2 where we have simulated fifty paths from the Lotka-Volterra SDE introduced in subsection 4.1 using the MDB of Durham and Gallant (2002). Again, for illustration purposes, paths have been plotted twice; the paths on the left have no transparency, whereas the

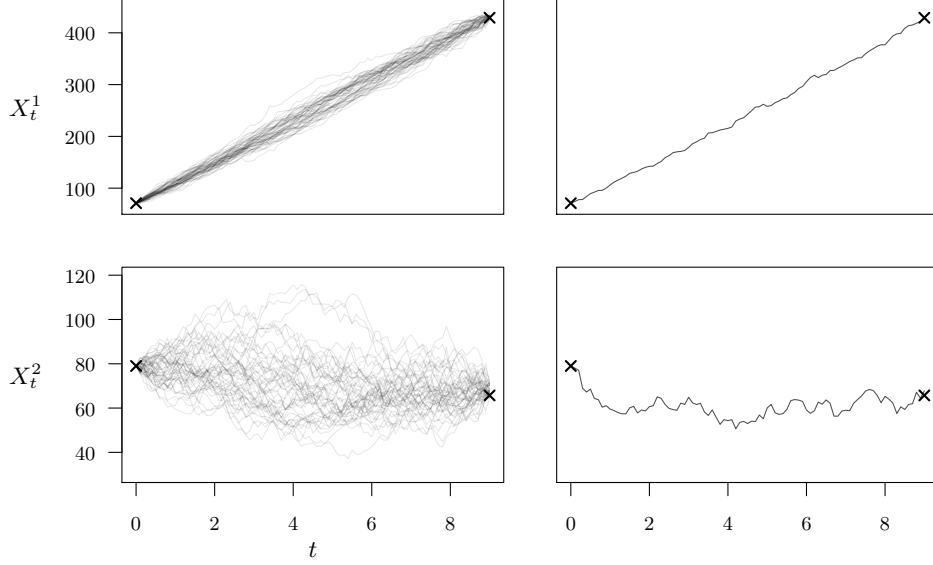


Figure 2: A plot of fifty paths simulated using the MDB of Durham and Gallant (2002) on the Lotka-Volterra SDE introduced in subsection 4.1. As in figure 1, the plots on the left are the fifty two-dimensional simulated paths with no transparency and the plots on the right are the fifty two-dimensional paths with transparency inversely proportional to their normalised weights. The two-dimensional initial condition and observation are illustrated with crosses. It can be seen that, even though all of the paths are consistent with the observation, none of the paths are consistent with the dynamics of the true diffusion and hence one of the paths has a much larger weight relative to the other paths.

paths on the right have transparency inversely proportional to their normalised weights.

Whitaker et al. (2016) introduce residual-bridge proposals which deal with this issue, albeit at a greater computational cost, by constructing a deterministic path, ξ_t , which captures the non-linear dynamics of the true, conditioned diffusion and considering the residual, $R_t := X_t - \xi_t$, which satisfies, for $t \in [0, T]$,

$$dR_t = (\mu(X_t, t) - \xi'_t)dt + \sigma(X_t, t)dW_t, \quad R_0 = 0.$$

If ξ_t accurately captures the non-linear dynamics of the true, conditioned diffusion then the residual should exhibit behaviour which is more linear, hence applying the MDB to the residual and adding back ξ_t will result in a proposal which more closely resembles the optimal proposal. Suppose, then, that at time t_k we have simulated x_k . Applying the MDB to the residual, R_t , gives the following joint distribution for the approximate residual process, R_t^{RB} , at the next point of the partition and at the observation time;

$$\begin{bmatrix} R_{k+1}^{\text{RB}} \\ R_K^{\text{RB}} \end{bmatrix} \bigg| (X_k^{\text{RB}} = x_k) \sim \text{N}(\gamma_k^{\text{RB}}, C_k^{\text{RB}}),$$

where

$$\begin{aligned}\gamma_k^{\text{RB}} &:= \begin{bmatrix} (x_k - \xi_k) + \Delta t(\mu(x_k, t_k) - (\xi_{k+1} - \xi_k)/\Delta t) \\ (x_k - \xi_k) + (T - t_k)(\mu(x_k, t_k) - (\xi_{k+1} - \xi_k)/\Delta t) \end{bmatrix} \\ &= \begin{bmatrix} (x_k - \xi_{k+1}) + \Delta t\mu(x_k, t_k) \\ (x_k - \xi_k) + (T - t_k)(\mu(x_k, t_k) - (\xi_{k+1} - \xi_k)/\Delta t) \end{bmatrix}, \\ C_k^{\text{RB}} &:= \begin{bmatrix} \Delta t\zeta(x_k, t_k) & \Delta t\zeta(x_k, t_k) \\ \Delta t\zeta(x_k, t_k) & (T - t_k)\zeta(x_k, t_k) \end{bmatrix}.\end{aligned}$$

Here $X_t^{\text{RB}} := R_t^{\text{RB}} + \xi_t$ denotes the process which approximates the true process and, as in Whitaker et al. (2016), we have approximated ξ'_k via the chord between (t_k, ξ_k) and (t_{k+1}, ξ_{k+1}) . Adding back ξ_t leads to the following joint distribution for the approximate process, X_t^{RB} , at the next point of the partition and at the observation time;

$$\begin{bmatrix} X_{k+1}^{\text{RB}} \\ X_K^{\text{RB}} \end{bmatrix} \Big| (X_k^{\text{RB}} = x_k) \sim \text{N} \left(\begin{bmatrix} m_{k+1}^{\text{RB}} \\ m_K^{\text{RB}} \end{bmatrix}, \Psi_k^{\text{RB}} \right),$$

where

$$\begin{aligned}m_{k+1}^{\text{RB}} &:= x_k + \Delta t\mu(x_k, t_k), \\ m_K^{\text{RB}} &:= x_k + (\xi_K - \xi_k) + (T - t_k)(\mu(x_k, t_k) - (\xi_{k+1} - \xi_k)/\Delta t), \\ \Psi_k^{\text{RB}} &:= \begin{bmatrix} \Delta t\zeta(x_k, t_k) & \Delta t\zeta(x_k, t_k) \\ \Delta t\zeta(x_k, t_k) & (T - t_k)\zeta(x_k, t_k) \end{bmatrix}.\end{aligned}$$

Therefore, the joint distribution for the approximate process at the next point of the partition and the observation, Y^1 , is given by

$$\begin{bmatrix} X_{k+1}^{\text{RB}} \\ Y^1 \end{bmatrix} \Big| (X_k^{\text{RB}} = x_k) \sim \text{N} \left(\begin{bmatrix} \bar{m}_{k+1}^{\text{RB}} \\ \bar{m}_K^{\text{RB}} \end{bmatrix}, \bar{\Psi}_k^{\text{RB}} \right),$$

where

$$\begin{aligned}\bar{m}_{k+1}^{\text{RB}} &:= x_k + \Delta t\mu(x_k, t_k), \\ \bar{m}_K^{\text{RB}} &:= P_1 x_k + P_1(\xi_K - \xi_k) + (T - t_k)P_1(\mu(x_k, t_k) - (\xi_{k+1} - \xi_k)/\Delta t), \\ \bar{\Psi}_k^{\text{RB}} &:= \begin{bmatrix} \Delta t\zeta(x_k, t_k) & \Delta t\zeta(x_k, t_k)P_1^* \\ \Delta tP_1\zeta(x_k, t_k) & (T - t_k)P_1\zeta(x_k, t_k)P_1^* + \Sigma_1 \end{bmatrix}.\end{aligned}$$

Standard manipulations for the multivariate normal distribution show that

$$(X_{k+1}^{\text{RB}} | X_k^{\text{RB}} = x_k, Y^1 = y^1) \sim \text{N}(a_k^{\text{RB}}, V_k^{\text{RB}}), \quad (2.4)$$

where

$$\begin{aligned}a_k^{\text{RB}} &:= x_k + \Delta t\mu(x_k, t_k) + \Delta t\zeta(x_k, t_k)P_1^*((T - t_k)P_1\zeta(x_k, t_k)P_1^* + \Sigma_1)^{-1} \\ &\quad \times (y_1 - P_1 x_k - P_1(\xi_K - \xi_k) - (T - t_k)P_1(\mu(x_k, t_k) - (\xi_{k+1} - \xi_k)/\Delta t)), \\ V_k^{\text{RB}} &:= \Delta t\zeta(x_k, t_k) - \Delta t^2\zeta(x_k, t_k)P_1^*((T - t_k)P_1\zeta(x_k, t_k)P_1^* + \Sigma_1)^{-1}P_1\zeta(x_k, t_k).\end{aligned}$$

The performance of such a proposal clearly hinges on choosing a deterministic path ξ_t which has similar dynamics to the true diffusion. One natural candidate² for ξ_t is constructed by ignoring the volatility in the true diffusion. That is, if we let $\xi_t \equiv \eta_t$ be the path obtained by ignoring any stochasticity in the evolution of the diffusion, then, from (1.2) we have that η_t satisfies

$$\eta_{t+\Delta t} = \eta_t + \Delta t \mu(\eta_t, t) + o(\Delta t) ,$$

for any $[t, t + \Delta t] \subset [0, T]$. Therefore, η_t solves the ordinary differential equation (ODE)

$$\frac{d\eta_t}{dt} = \mu(\eta_t, t) , \quad \eta_0 = x_0 , \quad (2.5)$$

over $[0, T]$. We denote the residual-bridge with this choice of ξ_t by RB^{ODE} . This choice for ξ_t is independent of the observation and hence can fail to capture the true dynamics of the *conditioned* diffusion, particularly when the noise in the observation, Σ_1 , is small and the difference between the observation, y^1 , and the endpoint of the deterministic path, η_K , is large. Therefore, paths simulated using this proposal can be inconsistent with the *conditioned* diffusion when the inter-observation time, T , is relatively large, since, for larger T , the stochasticity in the SDE results in dynamics which are inconsistent with η_t . As suggested by Whitaker et al. (2016), this motivates constructing a path ξ_t which is consistent with the *conditioned* diffusion by approximating the residual R_t with a tractable process \hat{R}_t and choosing

$$\xi_t = \eta_t + \mathbb{E}(\hat{R}_t | Y^1 = y^1) .$$

One choice³ for the tractable process \hat{R}_t is that given by the linear noise approximation (LNA). By Taylor expanding around η_t , defined by (2.5), the LNA constructs an \hat{R}_t which satisfies a linear SDE, and therefore has Gaussian transition densities. Indeed, by taking a first-order Taylor expansion of the drift and a zeroth-order Taylor expansion of the square-root of the volatility, one arrives at an approximate process \hat{R}_t which satisfies

$$d\hat{R}_t = J(\eta_t, t) \hat{R}_t dt + \sigma(\eta_t, t) dB_t , \quad \hat{R}_0 = 0 , \quad (2.6)$$

over the interval $[0, T]$, where $J(\eta_t, t)$ is the $d \times d$ Jacobian matrix whose (i, j) -th entry is

$$J(\eta_t, t)_{ij} := \left. \frac{\partial \mu(x, t)_i}{\partial x_j} \right|_{x=\eta_t} .$$

Under this approximation, a tractable form for $\mathbb{E}(\hat{R}_t | Y^1 = y^1)$ is available. The following lemma, whose proof is deferred to appendix A, derives a form which can be implemented in a computationally efficient manner because the ODEs that need to be solved do not involve any inverses.

²Justified for diffusions relating to the chemical Langevin equation by Theorem 2.1 in Chapter 11 of Ethier and Kurtz (1986).

³Justified for diffusions relating to the chemical Langevin equation by Theorem 2.3 in Chapter 11 of Ethier and Kurtz (1986)

Lemma 2.1. *Let \hat{R}_t be the process which satisfies (2.6) over the interval $[0, T]$ and let Y_1 be such that*

$$(Y^1 | \hat{R}_T = r) \sim N(P_1(r + \eta_T), \Sigma_1) .$$

Then

$$\mathbb{E}(\hat{R}_t | Y^1 = y^1) = \phi_t G_t^{-1} G_T^* P_1^* (P_1 \phi_T P_1^* + \Sigma_1)^{-1} (y_1 - P_1 \eta_T) ,$$

where G_t and ϕ_t satisfy, for $t \in [0, T]$, the following ODEs;

$$\begin{aligned} \frac{dG_t}{dt} &= J(\eta_t, t) G_t , & G_0 &= I , \\ \frac{d\phi_t}{dt} &= J(\eta_t, t) \phi_t + \phi_t J(\eta_t, t)^* + \zeta(\eta_t, t) , & \phi_0 &= 0 . \end{aligned}$$

For most diffusions G_t and ϕ_t will not be available analytically; however, using the Fortran subroutine `lsoda` (Petzold, 1983), both can be numerically evaluated in an accurate and efficient way at any point of the partition $\mathcal{P}_{\Delta t}^0$. We denote the residual-bridge proposal with this choice of ξ_t by RB^{LNA} .

Fifty paths simulated from the Lotka-Volterra SDE introduced in subsection 4.1 using the RB^{ODE} and RB^{LNA} proposals along with the corresponding deterministic paths can be seen in figures 3 and 4 respectively. As before, in both figures, the paths have been plotted twice; the paths on the left of each figure have no transparency, whereas the paths on the right of each figure have transparency inversely proportional to their normalised weights. Although such approaches account for the non-linear dynamics of the diffusion, they still assume a constant volatility over the region of interest. This leads to poor performance for diffusions whose volatility varies greatly over this interval and in particular, therefore, for larger inter-observation times T and for diffusions whose volatility is time-inhomogeneous.

3 New Proposals Based on Diffusion Bridges

We propose an extension to the approach of Whitaker et al. (2016) by constructing a *process*, U_t , which exhibits similar dynamics to the true, conditioned diffusion and considering the residual process $\tilde{R}_t := X_t - U_t$. We begin by constructing a deterministic path, ξ_t , which exhibits similar dynamics to the true diffusion (for instance, the path on which RB^{ODE} or RB^{LNA} is based). We then use this path to construct U_t which is coupled with the true diffusion through *the same driving Brownian motion* in such a way that paths of U_t exhibit similar stochastic behaviour to paths of X_t . Specifically, for an arbitrary u_0 , we define U_t to be the process which satisfies

$$dU_t = \xi'_t dt + \sigma(\xi_t, t) dB_t , \quad U_0 = u_0$$

over the interval $[0, T]$ and which is coupled with X_t through *the same driving Brownian motion*, B_t . The residual process, \tilde{R}_t , thus satisfies

$$d\tilde{R}_t = (\mu(X_t, t) - \xi'_t) dt + (\sigma(X_t, t) - \sigma(\xi_t, t)) dB_t ,$$

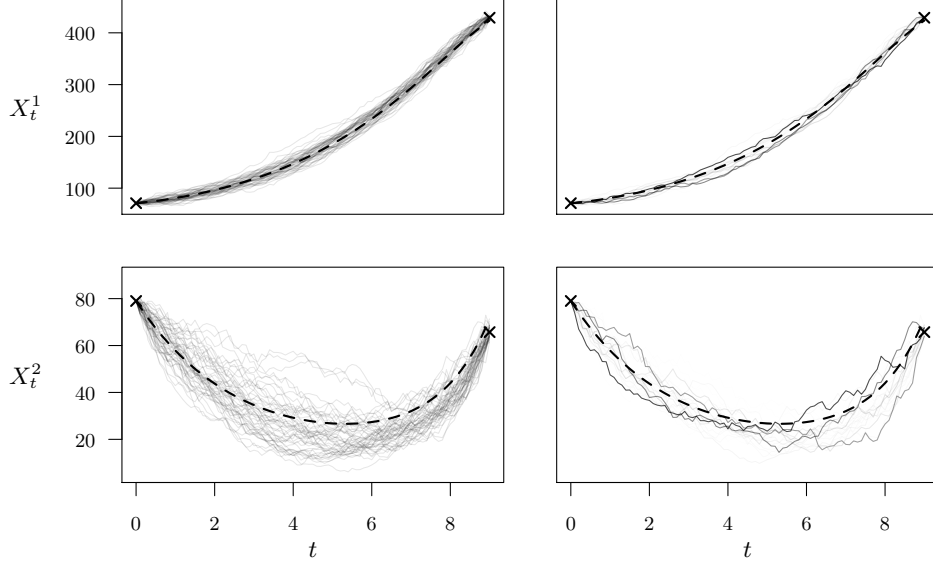


Figure 3: A plot of fifty paths simulated using the RB^{ODE} approach of Whitaker et al. (2016) from the Lotka-Volterra SDE introduced in subsection 4.1. As with the previous figures the plots on the left are the fifty two-dimensional simulated paths with no transparency and the plots on the right are the fifty two-dimensional paths with transparency inversely proportional to their normalised weights. The two-dimensional initial condition and observation are illustrated with crosses and the deterministic path, $\xi_t = \eta_t$ is plotted with a dashed line. When compared with previous figures these paths are more consistent with the true diffusion while still being consistent with the observation, thus the variability in the weights is smaller than the previous proposals.

over the interval $[0, T]$ and with initial condition $\tilde{R}_0 = x_0 - u_0$. We proceed by making the same approximation used in the MDB: suppose that we have simulated x_k at time t_k . Form an approximate process, \tilde{R}_t^{MDB} , which satisfies

$$d\tilde{R}_t^{\text{MDB}} = (\mu(x_k, t_k) - \xi'_{t_k}) dt + (\sigma(x_k, t_k) - \sigma(\xi_k, t_k)) dB_t ,$$

over the interval $[t_k, T]$ and has initial condition $\tilde{R}_k^{\text{MDB}} = x_k - u_k$ (where, as we shall see, u_k is the superfluous value of the process U_t at time t_k). With this approximation we have that, conditional on having simulated x_k at time t_k , the process $X_t^{\text{RB}} := U_t + \tilde{R}_t^{\text{MDB}}$ satisfies

$$dX_t^{\text{RB}} = (\xi'_t + \mu(x_k, t_k) - \xi'_{t_k}) dt + (\sigma(\xi_t, t) + \sigma(x_k, t_k) - \sigma(\xi_k, t_k)) dB_t ,$$

over the interval $[t_k, T]$ and has initial condition $X_k^{\text{RB}} = x_k$. Approximating σ by a piecewise constant function on the partition $\mathcal{P}_{\Delta t}$;

$$\sigma(\xi_u, u) = \sum_{k=0}^{K-1} \sigma(\xi_{t_k}, t_k) \mathbb{1}_{[t_k, t_{k+1})}(u) ,$$

where we denote by $\mathbb{1}_A(x)$ the indicator function on the set A , gives the following joint distribution for the approximate process, X^{RB} , at the next point of the partition and at

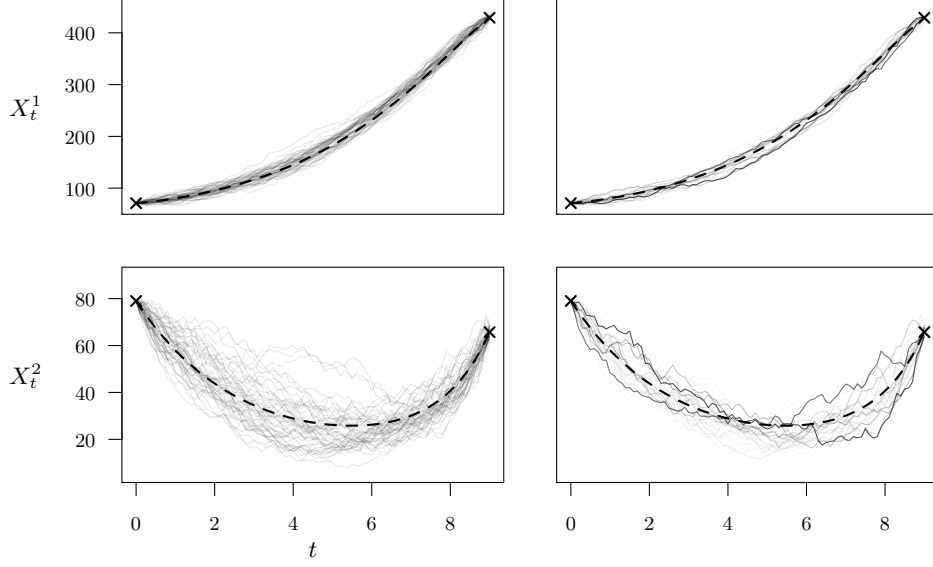


Figure 4: A plot of fifty paths simulated using the RB^{LNA} approach of Whitaker et al. (2016) from the Lotka-Volterra SDE introduced in subsection 4.1. As with the previous figures the plots on the left are the fifty two-dimensional simulated paths with no transparency and the plots on the right are the fifty two-dimensional paths with transparency inversely proportional to their normalised weights. The two-dimensional initial condition and observation are illustrated with crosses and the deterministic path, $\xi_t = \eta_t + \mathbb{E}(\hat{R}_t | Y^1 = y^1)$ is plotted with a dashed line. It can be seen here that, compared with figure 3, these paths are more consistent with the *conditioned* diffusion and therefore their corresponding weights are less variable.

the observation time;

$$\begin{bmatrix} X_{k+1}^{\text{RB}} \\ X_K^{\text{RB}} \end{bmatrix} \middle| (X_k^{\text{RB}} = x_k) \sim \text{N} \left(\begin{bmatrix} m_{k+1}^{\text{RB}} \\ m_K^{\text{RB}} \end{bmatrix}, \Psi_k^{\text{RB}} \right),$$

where

$$m_{k+1}^{\text{RB}} := x_k + \Delta t \mu(x_k, t_k),$$

$$m_K^{\text{RB}} := x_k + (\xi_K - \xi_k) + (T - t_k)(\mu(x_k, t_k) - (\xi_{k+1} - \xi_k)/\Delta t),$$

$$\Psi_k^{\text{RB}} := \begin{bmatrix} \Delta t \zeta(x_k, t_k) & \Delta t \zeta(x_k, t_k) \\ \Delta t \zeta(x_k, t_k) & \Psi_K^{\text{RB}} \end{bmatrix},$$

$$\Psi_K^{\text{RB}} := \Delta t \zeta(x_k, t_k) + \Delta t \sum_{j=k+1}^{K-1} [(\sigma(\xi_j, t_j) + \sigma(x_k, t_k) - \sigma(\xi_k, t_k))(\sigma(\xi_j, t_j) + \sigma(x_k, t_k) - \sigma(\xi_k, t_k))^*].$$

Thus, using (2.4), we see that

$$(X_{k+1}^{\text{RB}} | X_k^{\text{RB}} = x_k, Y_1 = y_1) \sim \text{N}(a_k^{\text{RB}}, V_k^{\text{RB}}), \quad (3.1)$$

where

$$\begin{aligned} a_k^{\overline{\text{RB}}} &:= x_k + \Delta t \mu(x_k, t_k) + \Delta t \zeta(x_k, t_k) P_1^* (P_1 \Psi_K^{\overline{\text{RB}}} P_1^* + \Sigma_1)^{-1} \\ &\quad \times (y_1 - P_1 x_k - P_1 (\xi_K - \xi_k) - (T - t_k) P_1 (\mu(x_k, t_k) - (\xi_{k+1} - \xi_k)/\Delta t)) , \\ V_k^{\overline{\text{RB}}} &:= \Delta t \zeta(x_k, t_k) - \Delta t^2 \zeta(x_k, t_k) P_1^* (P_1 \Psi_K^{\overline{\text{RB}}} P_1^* + \Sigma_1)^{-1} P_1 \zeta(x_k, t_k) . \end{aligned}$$

This proposal attempts to take into account the variability of the drift *and* the variability of the square-root of the volatility and therefore should outperform the previous residual-bridge construct in scenarios where the square-root of the volatility exhibits large variation over the interval $[0, T]$ and therefore, in particular, for relatively larger T and for volatilities which are time-inhomogeneous. A trade-off arises since if the square-root of the volatility varies too much then, in many cases of interest, constructing a deterministic path, ξ_t , which accurately captures the true dynamics of the diffusion will be tricky if not impossible. Moreover, in section 4.4 and in section 5 we highlight scenarios where our proposed bridge may be outperformed by the proposals of Whitaker et al. (2016). However, to illustrate why this new residual-bridge construct might be preferred over the residual-bridge construct of Whitaker et al. (2016), consider constructing bridges to the SDE

$$dX_t = \mu(t)dt + \sigma(t)dB_t , \quad X_0 = x_0 ,$$

over the interval $[0, T]$. It is clear that if one chooses ξ_t to be the solution of the ODE

$$\frac{d\xi_t}{dt} = \mu(t) , \quad \xi_0 = x_0 ,$$

then, for *any* $\sigma(t)$, this new proposal will, up to a discretisation error, simulate exact bridges of X_t , whereas, the proposal of Whitaker et al. (2016) will not. Moreover, the variability in the weights corresponding to the residual-bridge proposals of Whitaker et al. (2016) will increase the more $\sigma(t)$ varies over the region of interest.

As with the residual-bridge construct of Whitaker et al. (2016), ξ_t can be any deterministic path whose dynamics closely match that of the true conditioned diffusion. We denote this new proposal, where $\xi_t = \eta_t$ with η_t defined by (2.5), by $\overline{\text{RB}}^{\text{ODE}}$ and, where $\xi_t = \eta_t + \mathbb{E}(\hat{R}_t | Y_1 = y_1)$ with \hat{R}_t defined by (2.6), by $\overline{\text{RB}}^{\text{LNA}}$. Paths simulated using this proposal look very similar to paths simulated using the residual bridge proposals of Whitaker et al. (2016) as can be seen by comparing figures 3 and 4, with figures 5 and 6 which show fifty paths simulated from the Lotka-Volterra SDE introduced in subsection 4.1 using the $\overline{\text{RB}}^{\text{ODE}}$ and $\overline{\text{RB}}^{\text{LNA}}$ approaches respectively along with the corresponding deterministic paths, ξ_t . As throughout this paper, in both figures, the paths have been plotted twice; the paths on the left of each figure have no transparency, whereas the paths on the right of each figure have transparency inversely proportional to their normalised weights. By comparing the plots on the right of each figure with the corresponding plots on the right of figures 3 and 4 it can be seen that paths simulated using this proposal are more consistent with the true *conditioned* diffusion and thus have less variable weights.

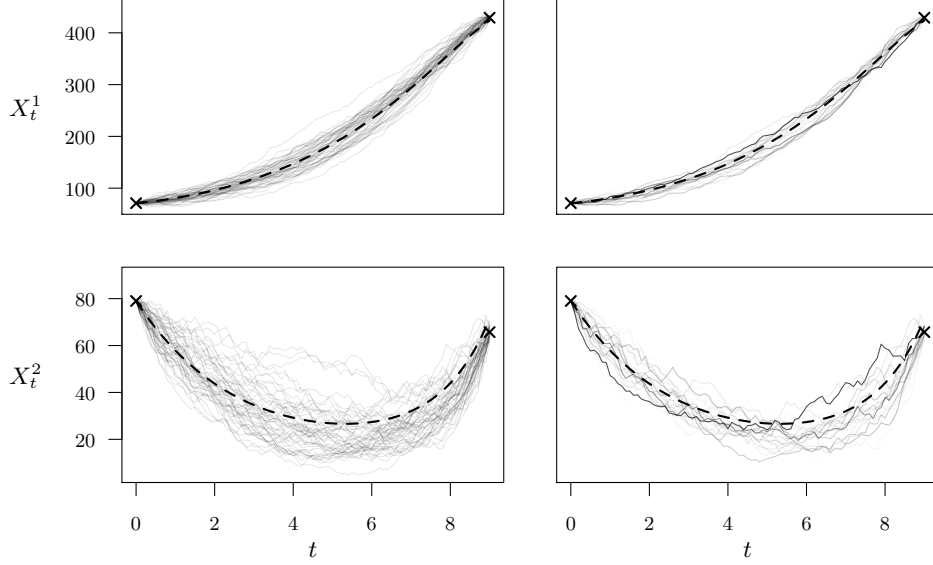


Figure 5: A plot of fifty paths simulated using the $\overline{\text{RB}}^{\text{ODE}}$ approach introduced in this paper from the Lotka-Volterra SDE introduced in subsection 4.1. As with the previous figures the plots on the left are the fifty two-dimensional simulated paths with no transparency and the plots on the right are the fifty two-dimensional paths with transparency inversely proportional to their normalised weights. The two-dimensional initial condition and observation are illustrated with crosses and the deterministic path, $\xi_t = \eta_t$ is plotted with a dashed line. When compared with figure 3 these paths are more consistent with the true *conditioned* diffusion thus the variability in the weights is smaller.

3.1 Computational Considerations

Comparing the form of $\Psi_k^{\overline{\text{RB}}}$ with the form of Ψ_k^{RB} , it can be seen that the residual-bridge proposals introduced in this paper have a larger computational cost compared to the corresponding residual-bridge proposals of Whitaker et al. (2016). We point out, however, that this difference in cost can be considerably reduced for diffusions relating to the chemical Langevin diffusion (see section 1 and the references therein), where the volatility is of the form

$$\zeta(x, t) = S\Lambda(x, t)^2 S^*,$$

where $S \in \mathbb{R}^{d \times r}$ is a constant matrix, and $\Lambda \in \mathbb{R}^{r \times r}$ is a diagonal matrix. In this case we can circumvent the calculation of partial sums of symmetric matrices of size $d \times d$ involved in the calculation of $\Psi_K^{\overline{\text{RB}}}$ and instead calculate partial sums of vectors of size r by letting $\sigma(x, t) = S\Lambda(x, t)$ so that

$$\Psi_K^{\overline{\text{RB}}} := \Delta t \zeta(x_k, t_k) + S \Delta t \sum_{j=k+1}^{K-1} [(\Lambda(\xi_j, t_j) + \Lambda(x_k, t_k) - \Lambda(\xi_k, t_k))(\Lambda(\xi_j, t_j) + \Lambda(x_k, t_k) - \Lambda(\xi_k, t_k))^*] S^*.$$

Thus, if r is significantly smaller than $d^2/2$, the computational cost of calculating $\Psi_k^{\overline{\text{RB}}}$ can be significantly reduced.

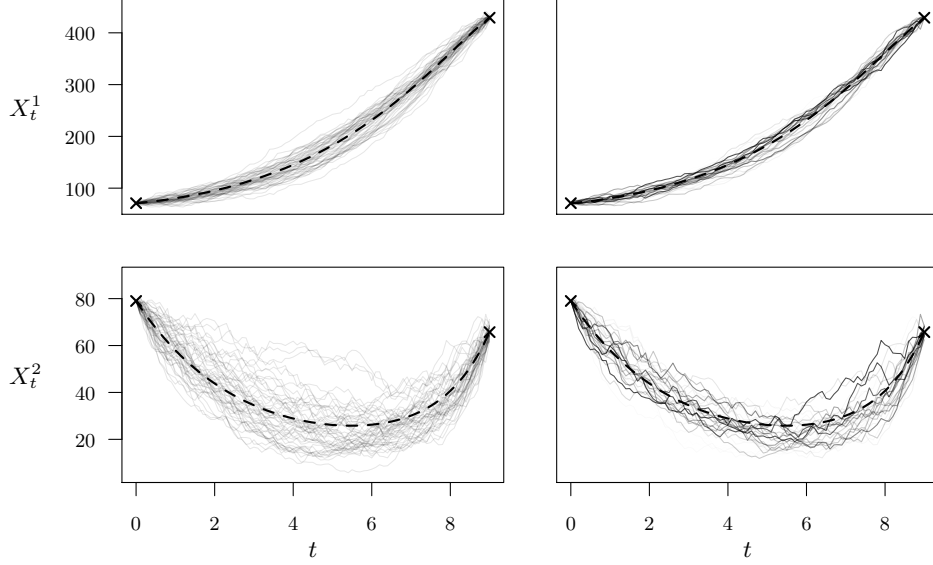


Figure 6: A plot of fifty paths simulated using the $\overline{\text{RB}}^{\text{LNA}}$ approach introduced in this paper from the Lotka-Volterra SDE introduced in subsection 4.1. As with the previous figures the plots on the left are the fifty two-dimensional simulated paths with no transparency and the plots on the right are the fifty two-dimensional paths with transparency inversely proportional to their normalised weights. The two-dimensional initial condition and observation are illustrated with crosses and the deterministic path, $\xi_t = \eta_t + \mathbb{E}(\hat{R}_t | Y^1 = y^1)$ is plotted with a dashed line. When compared with figure 4 these paths are more consistent with the true *conditioned* diffusion thus the variability in the weights is smaller.

4 A Simulation Study

In this section we compare the performance of the residual-bridge constructs introduced in this paper against the corresponding residual-bridge constructs of Whitaker et al. (2016) and the MDB construct of Durham and Gallant (2002) on two diffusions; the Lotka-Volterra (LV) diffusion (4.1) and a diffusion corresponding to a simple model of gene expression (GE, 4.2).

4.1 The Lotka-Volterra Diffusion

The Lotka-Volterra diffusion (e.g. Wilkinson, 2011) is an approximate model for the evolution of the numbers, $X_t = [X_t^1, X_t^2]^*$ of two species (prey and predators respectively) which are subject to three forces; prey reproduce with rate θ_1 , predators reproduce through eating prey with rate θ_2 , and predators die with rate θ_3 . Such a diffusion satisfies

$$\begin{bmatrix} dX_t^1 \\ dX_t^2 \end{bmatrix} = \begin{bmatrix} \theta_1 X_t^1 - \theta_2 X_t^1 X_t^2 \\ \theta_2 X_t^1 X_t^2 - \theta_3 X_t^2 \end{bmatrix} dt + \begin{bmatrix} \theta_1 X_t^1 + \theta_2 X_t^1 X_t^2 & -\theta_2 X_t^1 X_t^2 \\ -\theta_2 X_t^1 X_t^2 & \theta_2 X_t^1 X_t^2 + \theta_3 X_t^2 \end{bmatrix}^{1/2} dB_t ,$$

where, for a matrix A , $A^{1/2}$ denotes any matrix square-root so that $(A^{1/2})(A^{1/2})^* = A$.

4.2 A Diffusion for a Simple Gene Expression Model

In this subsection we introduce the diffusion which approximates a simple model for gene expression (see, for example, Komorowski et al., 2009; Golightly et al., 2015). This diffusion approximately describes the evolution of the numbers, $X_t = [R_t, P_t]^*$ of two biochemical species (mRNA and protein molecules respectively) which are subject to three forces; transcription with a time-inhomogeneous rate $k_R(t)$, mRNA degradation with rate γ_R , translation with rate k_P , and protein degradation with rate γ_P . As in Komorowski et al. (2009); Golightly et al. (2015) we take the rate $k_R(t)$ to be of the form

$$k_R(t) = b_0 \exp(-b_1(t - b_2)^2) + b_3 ,$$

so that the complete vector of unknown parameters is

$$\theta = (\gamma_R, \gamma_P, k_P, b_0, b_1, b_2, b_3) .$$

Such a diffusion satisfies

$$\begin{bmatrix} dR_t \\ dP_t \end{bmatrix} = \begin{bmatrix} k_R(t) - \gamma_R R_t \\ k_P R_t - \gamma_P P_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{k_R(t) + \gamma_R R_t} & 0 \\ 0 & \sqrt{k_P R_t + \gamma_P P_t} \end{bmatrix} dB_t .$$

We use the same parameters, θ , and initial conditions, x_0 , as those used in Whitaker et al. (2016) for the Lotka-Volterra diffusion;

$$\theta = (\theta_1, \theta_2, \theta_3) = (0.5, 0.0025, 0.3) , \quad x_0 = (71, 79) ,$$

and we use the following parameters,

$$\theta = (\gamma_R, \gamma_P, k_P, b_0, b_1, b_2, b_3) = (0.7, 0.72, 3, 80, 0.05, 2, 50) ,$$

and initial condition $x_0 = (70, 70)$ for the diffusion corresponding to the simple model of gene expression. We fix Δt to be 0.1 for the LV diffusion and 0.01 for the GE diffusion and choose 10 equally-spaced values for T between; 0 and 10 for the LV diffusion and 0 and 4 for the GE diffusion. Moreover, to compare the performance of the proposals in challenging scenarios, we choose $P_1 = I$ and $\Sigma_1 = 10^{-12}I$ so that the observation;

$$Y^1 | X_K = x \sim N(x, 10^{-12}I) ,$$

essentially corresponds to exact observations of the diffusion⁴ For each value of T , we simulated 10,000 values for Y_T^1 (where we have emphasised the dependence on T) using

⁴This small choice of variance in the observation is purely to generate challenging scenarios. In practice, if exact observations of the diffusion were available, the inference procedure would be slightly different (see, for example, Pedersen, 1995; Durham and Gallant, 2002) and is considered not here.

the EM approximation to forward simulate values of the path at points of the partition. For each collection of 10,000 values we chose five terminal points for y_T^1 , corresponding to the mean, along with the four 90% quantiles along the axes of the principal components. For each combination of (T, y_T^1) , we ran the MDB of Durham and Gallant (2002), the residual-bridge construct of Whitaker et al. (2016) with the two choices for ξ_t , RB^{ODE} and RB^{LNA} , along with the residual-bridge construct introduced in this paper with the same two choices for ξ_t , $\overline{\text{RB}}^{\text{ODE}}$ and $\overline{\text{RB}}^{\text{LNA}}$. For each of the five constructs, we simulated $N = 1,000,000$ independent skeleton paths and calculated the effective sample size per second (ESS/s) from the normalised importance weights (Liu, 1996):

$$\text{ESS/s } (\tilde{w}_{1:N}) = \frac{(\tilde{w}_1^2 + \dots + \tilde{w}_N^2)^{-1}}{\text{execution time}} . \quad (4.1)$$

To account for variability in the execution time, we calculated the average execution time over ten identical runs.

For completeness we have included, in appendix D, the relative effective sample sizes defined by

$$\text{Rel. ESS } (\tilde{w}_{1:N}) = N^{-1}(\tilde{w}_1^2 + \dots + \tilde{w}_N^2)^{-1} , \quad (4.2)$$

along with the (average) execution times for each proposal and for each combination of (T, y_1^T) for the Lotka-Volterra and gene-expression diffusions detailed in this section, and the birth-death diffusion detailed in appendix B.

4.3 Results

To ease visualisation of comparative performance, figures 7 and 8, which illustrate the results for the LV and GE diffusion respectively, plot, for four pairs of proposals, the effective sample size per second for one of the pair of proposals relative to the other for each combination of (T, y_T^1) for which both proposals had an effective sample size of at least one hundred. The four pairs of proposals are chosen to approximate the sequential ordering in which the paper has been presented. We emphasise that the larger the ESS/s the more statistically efficient the proposal is for that particular choice of inter-observation time T and observation y_T^1 . These figures illustrate that the effective sample size per second of the residual-bridge construct introduced in this paper is often similar to or larger than the effective sample size per second of the corresponding (in the sense of the same deterministic path) residual-bridges constructs of Whitaker et al. (2016) and, for larger inter-observation times, T , can exceed it by several orders of magnitude.

4.4 Issues Surrounding Robustness

Preserving the discrepancy between the square-root volatility at any particular time t_k , $\sigma(x_k, t_k)$, and the square-root volatility at the same time evaluated at the approximating

deterministic path, $\sigma(\xi_k, t_k)$, can be detrimental to the performance of the new residual-bridge proposals when compared to the proposals of Whitaker et al. (2016) in scenarios where preserving such a discrepancy leads to a large overestimate/underestimate of the true integrated volatility. Therefore it can be argued that this new residual-bridge proposal is less robust than the residual-bridge proposals of Whitaker et al. (2016) and care must be taken when implementing such a proposal. We illustrate this lack of robustness in appendix B where we compare the performance of the proposals on a simple one-dimensional diffusion.

4.5 Issues Surrounding Absolute Continuity

In this paper we do not prove that the limiting processes (as $\Delta t \downarrow 0$) corresponding to the proposals introduced in this paper are absolutely continuous with respect to the true conditioned diffusion. Therefore, even though decreasing Δt will decrease the bias in our approximate inference scheme, this decrease may come at an ever increasing variance, as measured by the variability in the weights (i.e. the effective sample size). However, we provide, through a further simulation study detailed in appendix C, numerical evidence suggesting that our proposals are robust to a decreasing Δt . Specifically, we look at the relative effective sample size (4.2) for the new residual-bridge proposals introduced in this paper for a variety of inter-observation times T , observations y_1^T , and step-sizes Δt when applied to three diffusions; the Lotka-Volterra diffusion and the diffusion for a simple model for gene expression introduced in this section, along with the birth-death diffusion introduced in appendix B. We illustrate that the relative effective sample sizes are consistent for

$$\Delta t = 0.1, 0.05, 0.01, 0.005, 0.001 ,$$

for the birth-death and gene expression diffusions, and for

$$\Delta t = 0.01, 0.005, 0.001, 0.0005, 0.0001 ,$$

for the Lotka-Volterra diffusion. This consistency of relative effective sample sizes, particularly over small values of Δt (which we emphasise are values of Δt that border on what is computationally feasible), strongly suggests that the proposals introduced in this paper can be implemented in any computationally feasible algorithm (i.e. one that does not use a prohibitively small Δt) without worrying about the effect that decreasing Δt has on the variability of the resulting weights.

5 Discussion of Results

The results of the simulation study illustrated in figures 7 and 8 show that the performance, in terms of the effective sample size per second, of the residual-bridge proposal introduced in this paper is often similar to or larger than the performance of the residual-bridge proposals of Whitaker et al. (2016), and, for larger inter-observation times, T , can

exceed it by several orders of magnitude. Therefore, when looked at in conjunction with the analysis of Whitaker et al. (2016), a particle MCMC scheme which uses this proposal will be more efficient than a particle MCMC scheme which uses any existing proposal and the potential gains in efficiency are large. However, as we highlight in appendix B, there exist some instances where these new residual-bridge constructs can have a lower (by a factor of one half in the worst case found in our simulation study) effective sample size per second than the corresponding constructs of Whitaker et al. (2016). Indeed, one drawback of the proposed residual-bridge constructs stems from the fact that, at intermediate time points, discrepancies of sample paths of the conditional diffusion from the deterministic path, ξ_t , can be relatively large; preserving the resulting discrepancies in the drift and volatility, when for $\overline{\text{RB}}^{\text{LNA}}$ these should be 0 at time T , for example, must be sub-optimal. An interpolation scheme which is both *justifiable* and *computationally efficient*, however, eludes us.

In this paper we have motivated the need for the construction of efficient proposals for approximately simulating conditioned diffusions over an interval $[0, T]$. We have briefly described some of the current proposals used in the literature and their drawbacks. We have introduced a new residual-bridge proposal and have explained, and demonstrated numerically, that such a proposal can often lead to larger effective sample sizes for a fixed computational budget, particularly for larger inter-observation times T and for diffusions with volatilities which are time-inhomogeneous. We have also highlighted, via a simulation study on a simple one-dimensional diffusion, that care needs to be taken when using such proposals as they are arguably less robust, across different diffusions, than the residual-bridge proposals of Whitaker et al. (2016). Further, we have provided numerical evidence which suggests that these new proposals are robust to a decreasing step-size Δt .

All the algorithms in this paper were written in modern Fortran, compiled using GNU Fortran (version 4.8.4) from the GNU Compiler Collection (<http://gcc.gnu.org/>) and implemented on an Intel Xeon E5-2699 v3 CPU.

Acknowledgements

S. Malory gratefully acknowledges the support of the EPSRC funded EP/H023151/1 STOR-i centre for doctoral training.

References

- Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation for stochastic volatility models. *Journal of Financial Economics*, 83(413). 1
- Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342. 2, 5

- Beskos, A., Papaspiliopoulos, O., and Roberts, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098. 2, 4
- Coffey, W., Kalmykov, Y., and Waldron, J. (2004). *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry, and Electrical Engineering*. Series in contemporary chemical physics. World Scientific. 1
- Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–338. 1, 2, 6, 7, 8, 16, 17, 18, 23, 30
- Ethier, S. and Kurtz, T. (1986). *Markov processes: characterization and convergence*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley. 1, 10
- Fearnhead, P. (2008). Computational methods for complex stochastic systems: a review of some alternatives to mcmc. *Statistics and Computing*, 18(2):151–171. 3
- Fearnhead, P., Giagos, V., and Sherlock, C. (2014). Inference for reaction networks using the linear noise approximation. *Biometrics*, 70(2):457–466. 1
- Golightly, A., Henderson, D. A., and Sherlock, C. (2015). Delayed acceptance particle mcmc for exact inference in stochastic kinetic models. *Statistics and Computing*, 25(5):1039–1055. 17
- Golightly, A. and Wilkinson, D. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis*, 52(3):1674 – 1693. 5
- Golightly, A. and Wilkinson, D. J. (2011). Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo. *Interface Focus*, 1(6):807–820. 1
- Komorowski, M., Finkenstädt, B., Harper, C. V., and Rand, D. A. (2009). Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, 10(1):1–10. 17
- Lindström, E. (2012). A regularized bridge sampler for sparsely sampled diffusions. *Statistics and Computing*, 22(2):615–623. 2, 3
- Liu, J. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119. 18
- Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):pp. 55–71. 2, 5, 6, 17
- Petzold, L. (1983). Automatic selection of methods for solving stiff and nonstiff sys-

tems of ordinary differential equations. *SIAM Journal on Scientific and Statistical Computing*, 4(1):136–148. 11

Rogers, L. and Williams, D. (2000). *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*. Cambridge Mathematical Library. Cambridge University Press. 3

Schauer, M., van der Meulen, F., and van Zanten, H. (2013). Guided proposals for simulating multi-dimensional diffusion bridges. *ArXiv e-prints*. 3

van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. 1

Whitaker, G. A., Golightly, A., Boys, R. J., and Sherlock, C. (2016). Improved bridge constructs for stochastic differential equations. *Statistics and Computing*, pages 1–16. 1, 3, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 23, 24, 30

Wilkinson, D. (2011). *Stochastic Modelling for Systems Biology, Second Edition*. Chapman & Hall/CRC Mathematical and Computational Biology. Taylor & Francis. 1, 16, 23

A A Proof of Lemma 2.1

Proof. Define the *generator*, G_t , as the solution to

$$\frac{dG_t}{dt} = J(\eta_t, t)G_t, \quad G_0 = I,$$

over the interval $[0, T]$. Consider the process $G_t^{-1}\hat{R}_t$ which satisfies

$$\begin{aligned} d(G_t^{-1}\hat{R}_t) &= dG_t^{-1}\hat{R}_t + G_t^{-1}d\hat{R}_t \\ &= -G_t^{-1}dG_tG_t^{-1}\hat{R}_t + G_t^{-1}J(\eta_t, t)\hat{R}_tdt + G_t^{-1}\sigma(\eta_t, t)dB_t \\ &= G_t^{-1}\sigma(\eta_t, t)dB_t. \end{aligned}$$

Therefore, for any $0 \leq s \leq t \leq T$, $G_t^{-1}\hat{R}_t$ is normally distributed with

$$\mathbb{E}(G_t^{-1}\hat{R}_t) = 0, \quad \text{Cov}(G_s^{-1}\hat{R}_s, G_t^{-1}\hat{R}_t) = \int_0^s G_u^{-1}\zeta(\eta_u, u)G_u^{-*} du,$$

where G^{-*} is shorthand for $(G^{-1})^*$. Let ψ_t be the solution to

$$\frac{d\psi_t}{dt} = G_t^{-1}\zeta(\eta_t, t)G_t^{-*}, \quad \psi_0 = 0, \tag{A.1}$$

over the interval $[0, T]$. Then

$$\begin{bmatrix} \hat{R}_t \\ Y_1 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ P_1\eta_T \end{bmatrix}, \begin{bmatrix} G_t\psi_tG_t^* & G_t\psi_tG_T^*P_1^* \\ P_1G_T\psi_tG_t^* & P_1G_T\psi_TG_T^*P_1^* + \Sigma_1 \end{bmatrix}\right).$$

Therefore

$$\mathbb{E}(\hat{R}_t|Y_1 = y_1) = G_t\psi_tG_T^*P_1^*(P_1G_T\psi_TG_T^*P_1^* + \Sigma_1)^{-1}(y_1 - P_1\eta_T).$$

To circumvent the need to calculate ψ_t , and therefore avoid solving the costly ODE (A.1) which contains inverses on the right-hand side, we let $\phi_t := G_t \psi_t G_t^*$ and note that ϕ_t solves

$$\begin{aligned} \frac{d\phi_t}{dt} &= \frac{dG_t}{dt} \psi_t G_t^* + G_t \psi_t \frac{dG_t^*}{dt} + G_t \frac{d\psi_t}{dt} G_t^* \\ &= J(\eta_t, t) G_t \psi_t G_t^* + G_t \psi_t G_t^* J(\eta_t, t)^* + \zeta(\eta_t, t) \\ &= J(\eta_t, t) \phi_t + \phi_t J(\eta_t, t)^* + \zeta(\eta_t, t), \end{aligned}$$

over the interval $[0, T]$ with initial condition $\phi_0 = 0$. □

B Issues Surrounding Robustness

In this appendix we illustrate, via a simulation study, that the new residual-bridge constructs introduced in this paper can have a lower effective sample size per second than the residual bridge constructs of Whitaker et al. (2016) and are arguably less robust over different diffusions. We will consider a one-dimensional, birth-death diffusion X_t (Wilkinson, 2011) which satisfies

$$dX_t = (\theta_1 - \theta_2)X_t dt + \sqrt{(\theta_1 + \theta_2)X_t} dB_t, \quad X_0 = x_0$$

over the interval $[0, T]$. This diffusion can be considered as an approximate model for the evolution of the number, X_t , of a species which is subject to two forces; births and deaths with rates θ_1 and θ_2 respectively. Due to the simplicity of the drift and volatility of this diffusion, the term η_t , defined by (2.5), along with the terms G_t and ϕ_t defined in lemma 2.1 are analytically tractable with $\eta_t = x_0 \exp((\theta_1 - \theta_2)t)$, $G_t = \exp((\theta_1 - \theta_2)t)$, and

$$\phi_t = \frac{(\theta_1 + \theta_2)}{(\theta_1 - \theta_2)} \eta_t (\exp((\theta_1 - \theta_2)t) - 1).$$

We conduct a simulation study which mimics the simulation study of section 4 in order compare the performance of the residual-bridge construct introduced in this paper against the residual-bridge construct of Whitaker et al. (2016) and the MDB construct of Durham and Gallant (2002) on the birth-death diffusion. We use the same parameters, θ , and initial conditions, x_0 , as those used in Whitaker et al. (2016); $(\theta_1, \theta_2) = (0.1, 0.8)$, $x_0 = 50$, so that sample paths of the diffusion exhibit exponential decay. We fix Δt to be 0.01 and choose 10 equally-spaced values for T between 0 and 2. Moreover, we choose $P_1 = 1$ and $\Sigma_1 = 10^{-12}$ so that the observation;

$$Y^1 | X_K = x \sim N(x, 10^{-12}),$$

essentially corresponds to exact observations of the diffusion. For each value for T , we simulated 10,000 values for Y_T^1 (where we have emphasised the dependence on T) using the EM approximation to forward simulate values of the path at points of the partition. For each collection of 10,000 values we chose three terminal points for y_T^1 , corresponding to the 5%, 50%, and 95% quantiles. For each combination of (T, y_T^1) , we ran the MDB of Durham and Gallant (2002), the residual-bridge construct of Whitaker et al. (2016) with the two choices for ξ_t , RB^{ODE} and RB^{LNA} , along with the residual-bridge construct introduced in this paper with the same two choices for ξ_t , $\overline{\text{RB}}^{\text{ODE}}$ and $\overline{\text{RB}}^{\text{LNA}}$. For each of the five constructs, we simulated $N = 1,000,000$ independent skeleton paths and calculated, from the normalised importance weights, the effective sample size per second (ESS/s) as defined by (4.1)⁵. As previously, to ease visualisation of

⁵As before, to mitigate variability in the execution time, we calculated the average execution time over ten identical runs.

comparative performance, figure 9 plots, for four pairs of proposals, the effective sample size per second for one of the pair of proposals relative to the other for each combination of (T, y_T^1) . The four pairs of proposals are chosen to approximate the sequential ordering in which the paper has been presented. Again, we emphasise that the larger the ESS/s the more statistically efficient the proposal is for that particular choice of inter-observation time T and observation y_T^1 .

Figure 9 illustrates that the effective sample size per second for the new residual-bridge construct which uses η_t , defined by (2.5), as the deterministic path is similar to, but slightly smaller than, due to the increase in computational cost, the effective sample size per second for the corresponding residual-bridge construct of Whitaker et al. (2016). However, the new residual-bridge construct which uses $\eta_t + \mathbb{E}(\hat{R}_t|Y_1 = y_1)$, with η_t defined by (2.5) and \hat{R}_t defined by (2.6), as the deterministic path has an effective sample size per second which is significantly worse than the corresponding residual-bridge construct of Whitaker et al. (2016) for the observations corresponding to the 5% and 95% quantiles. This difference is particularly large for the observation corresponding to the 5% quantile and demonstrates that the performance of the new residual-bridge proposals introduced in this paper can be worse than that of the residual-bridge proposals of Whitaker et al. (2016), thus care needs to be taken when implementing such a proposal. We note that, in this example, one can transform the diffusion to a diffusion with unit volatility. Specifically, if we let

$$Y_t := 2\sqrt{\frac{X_t}{(\theta_1 + \theta_2)}},$$

then Y_t satisfies

$$dY_t = \left(\frac{(\theta_1 - \theta_2)}{2} Y_t - \frac{1}{2Y_t} \right) dt + dB_t, \quad Y_0 = 2\sqrt{\frac{x_0}{(\theta_1 + \theta_2)}}.$$

As the volatility is constant, applying the residual-bridge construct introduced in this paper to the transformed diffusion is equivalent to applying the residual-bridge construct of Whitaker et al. (2016) to the transformed diffusion and thus the resulting effective sample sizes will be identical (provided, of course, the same random numbers are used). However, we emphasise that in most cases of practical interest one will not be able to transform the diffusion to one of unit volatility and therefore care *must* be taken when implementing the residual-bridge constructs introduced in this paper.

For completeness we have included, in appendix D, the relative effective sample sizes (as defined by (4.2)) along with the execution times for each proposal and for each combination of (T, y_1^T) for the birth-death diffusion detailed in this appendix, and for the Lotka-Volterra and gene-expression diffusions detailed in section 4.

C Issues Surrounding Absolute Continuity

In this appendix we provide numerical evidence, via a simulation study, suggesting that the residual-bridge proposals introduced in this paper are robust to a decreasing step-size, Δt . This simulation study will partially extend the studies in section 4 and appendix B by considering the two residual-bridge constructs introduced in this paper; $\overline{\text{RB}}^{\text{ODE}}$ and $\overline{\text{RB}}^{\text{LNA}}$, three diffusions; birth-death, Lotka-Volterra, and a diffusion corresponding to a simple model of gene expression, and using the same parameters and initial conditions as those used in section 4 and appendix B. To test the proposals in a broad variety of scenarios we chose three values for T ; $(0.2, 1, 2)$ for the

BD diffusion, (1, 4, 7) for the LV diffusion, and (0.4, 2, 3.6) for the GE diffusion, corresponding to a small, medium and large inter-observation interval. For each value of T we chose two observations y_1^T from the set of observations simulated for the simulation studies in section 4 and appendix B; the centre of the simulated observations and one other chosen at random. We chose five different values for Δt ; (0.01, 0.005, 0.001, 0.0005, 0.0001) for the BD and GE diffusions and (0.1, 0.05, 0.01, 0.005, 0.001) for the LV diffusion. For each proposal and each combination of $(T, y_1^T, \Delta t)$ we simulated $N = 1,000,000$ independent skeleton paths and calculated the relative effective sample size (as defined by (4.2)) from the normalised importance weights⁶.

⁶For all of the models and observations the observation variance that was used, 10^{-12} , is several orders of magnitude smaller than the eigenvalues of the variance matrix at the observation so the empirical evidence of absolute continuity is not affected by this.

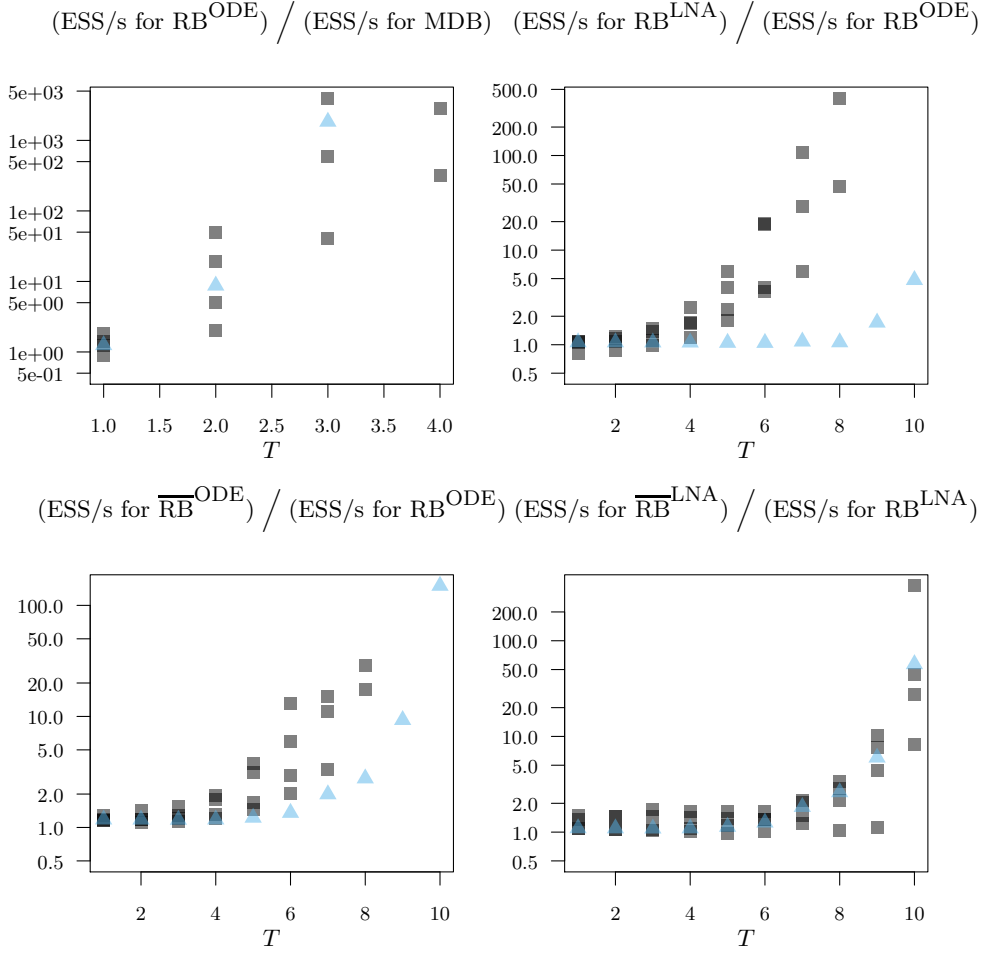


Figure 7: Plots of the comparative effective sample size per second for four pairs of proposals and for a variety of combinations of (T, y_1^T) corresponding to the Lotka-Volterra diffusion. Observations, y_1^T , corresponding to the mean of the simulated observations are denoted with blue triangles, whereas the observations corresponding to the four 90% quantiles along the axes of the principal components are denoted with grey boxes.

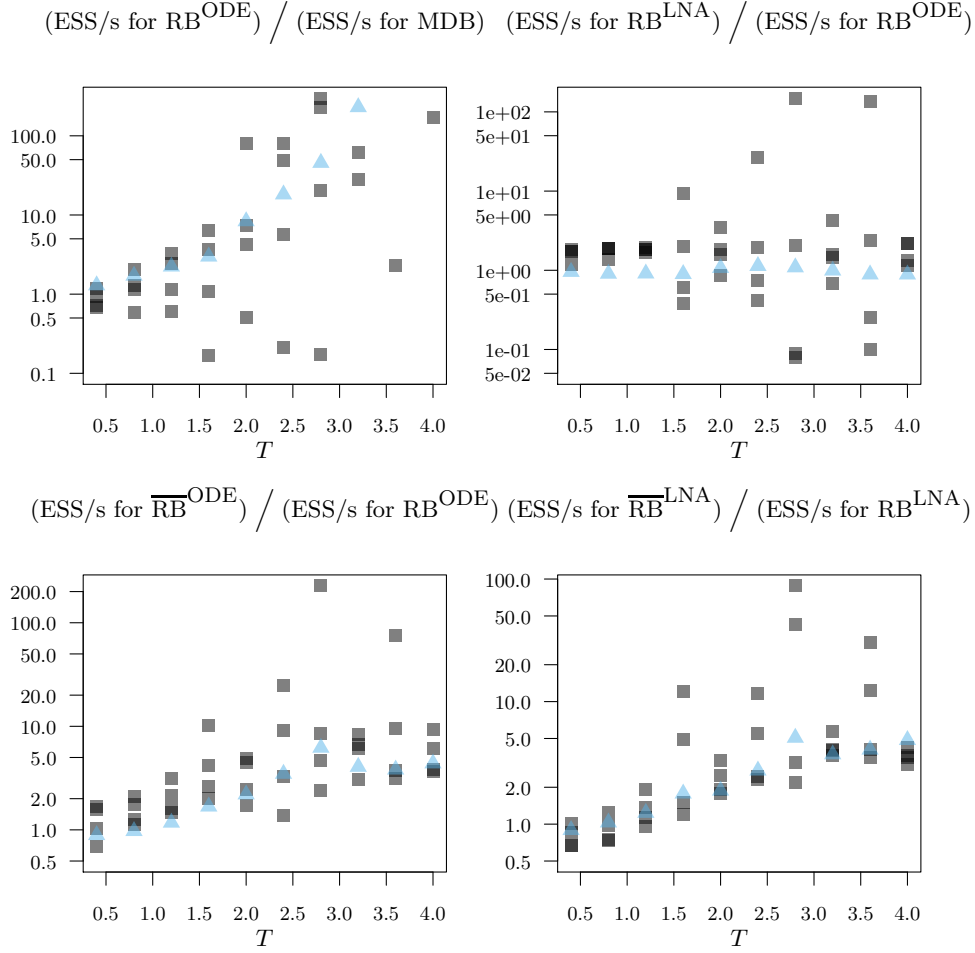


Figure 8: Plots of the comparative effective sample size per second for four pairs of proposals and for a variety of combinations of (T, y_1^T) corresponding to the Lotka-Volterra diffusion. The format is the same as figure 8.

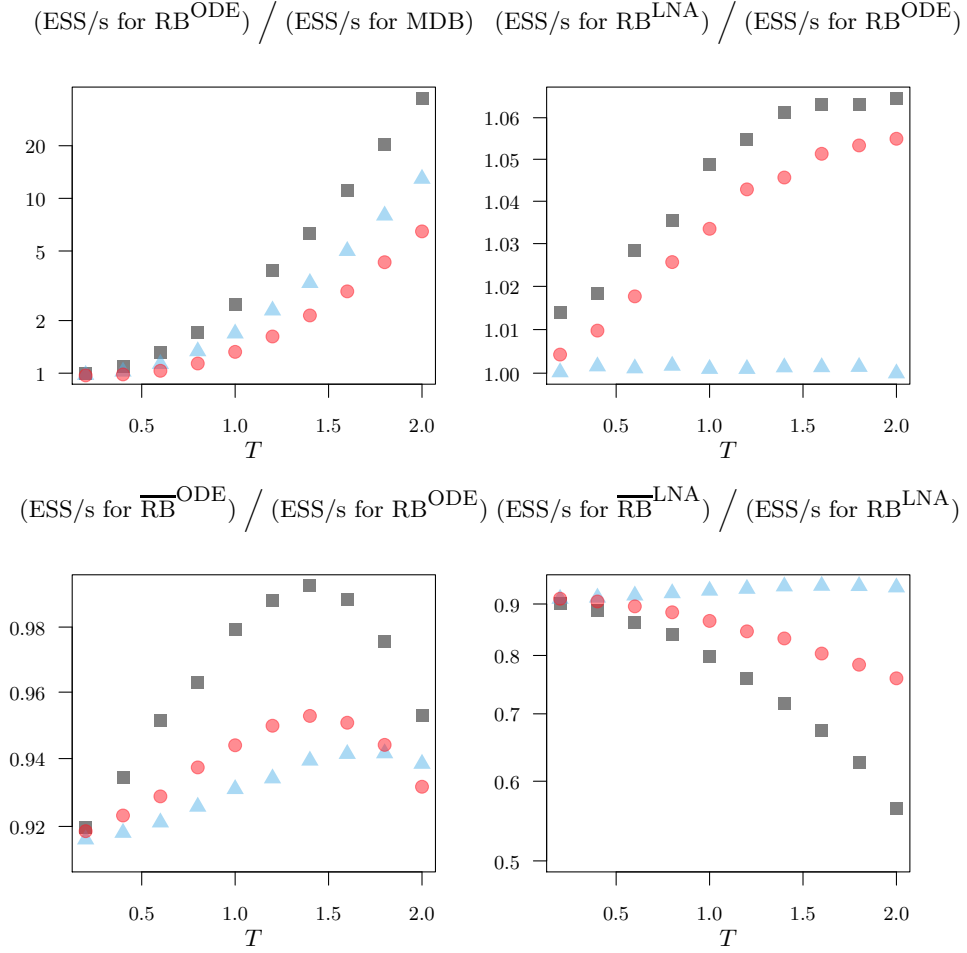


Figure 9: Plots of the comparative effective sample size per second for four pairs of proposals and for a variety of combinations of (T, y_1^T) corresponding to the birth-death diffusion. The observations, y_1^T , corresponding to the 50% quantile of the simulated observations are denoted with blue triangles, the observations corresponding to the 5% quantile are denoted with grey boxes, and the observations corresponding to the 95% quantile are denoted with red circles.

Table 1: A table showing the relative effective sample sizes for 1,000,000 independent skeleton paths simulated from the two proposals; $\overline{\text{RB}}^{\text{ODE}}$ and $\overline{\text{RB}}^{\text{LNA}}$ for a variety of diffusion models (birth-death, Lotka-Volterra, and gene-expression), inter-observation times (small, medium, and large), step-sizes, and observations (the centre, and one other chosen at random for each combination of (model, T, y_1^T), but fixed for the different step-sizes). The range of step-sizes are $\Delta t = 0.01, 0.005, 0.001, 0.0005, 0.0001$ for the BD and GE diffusions, and $\Delta t = 0.1, 0.05, 0.01, 0.005, 0.001$ for the LV diffusion and the results are displayed in decreasing step-size order. That is, for each group of five results, corresponding to the different values for Δt , the effective sample size corresponding to the largest and smallest value for Δt is at the top and bottom of the group respectively.

Proposal		$\overline{\text{RB}}^{\text{ODE}}$						$\overline{\text{RB}}^{\text{LNA}}$				
Diffusion Model	Birth-Death		Lotka-Volterra		Gene-Expression		Birth-Death		Lotka-Volterra		Gene-Expression	
Observation	Centre	Other	Centre	Other	Centre	Other	Centre	Other	Centre	Other	Centre	Other
Small T	0.9992	0.9990	0.9719	0.9350	0.9370	0.8407	0.9992	0.9986	0.9716	0.9621	0.9372	0.9054
	0.9995	0.9991	0.9733	0.9407	0.9408	0.8493	0.9995	0.9987	0.9731	0.9634	0.9409	0.9083
	0.9997	0.9992	0.9744	0.9449	0.9441	0.8568	0.9997	0.9987	0.9744	0.9643	0.9442	0.9107
	0.9997	0.9992	0.9745	0.9455	0.9444	0.8574	0.9997	0.9987	0.9745	0.9644	0.9445	0.9108
	0.9997	0.9992	0.9746	0.9460	0.9446	0.8581	0.9997	0.9987	0.9746	0.9644	0.9447	0.9112
Medium T	0.9926	0.9878	0.6635	0.4122	0.4289	0.2497	0.9925	0.9393	0.6574	0.6387	0.4289	0.4029
	0.9938	0.9890	0.6721	0.4396	0.4355	0.2263	0.9936	0.9408	0.6694	0.6514	0.4352	0.4118
	0.9947	0.9898	0.6767	0.4598	0.4469	0.2814	0.9944	0.9419	0.6765	0.6593	0.4468	0.4190
	0.9948	0.9899	0.6746	0.4566	0.4442	0.2721	0.9945	0.9421	0.6749	0.6582	0.4442	0.4107
	0.9948	0.9900	0.6755	0.4487	0.4372	0.2666	0.9946	0.9421	0.6757	0.6573	0.4370	0.4077
Large T	0.9367	0.9171	0.3379	0.0971	0.1404	0.0740	0.9344	0.7875	0.3350	0.3172	0.1403	0.1231
	0.9387	0.9208	0.3678	0.0839	0.1551	0.0806	0.9362	0.7926	0.3683	0.3289	0.1554	0.1401
	0.9405	0.9232	0.3688	0.0753	0.1557	0.0905	0.9378	0.7964	0.3756	0.3377	0.1556	0.1508
	0.9406	0.9230	0.3709	0.0743	0.1519	0.0785	0.9378	0.7968	0.3772	0.3381	0.1520	0.1502
	0.9406	0.9242	0.3664	0.0716	0.1624	0.0776	0.9378	0.7976	0.3727	0.3375	0.1628	0.1227

Table 1 shows that the relative effective sample size for the proposals introduced in this paper are consistent across varying values of Δt for the scenarios considered in the simulation study. This therefore suggests that such proposals can be implemented without the need to consider the effect that decreasing the step-size, Δt , has on the resulting variability of the weights. Moreover, we stress that the smallest Δt considered here is on the border of what is computationally feasible, in the sense that any smaller Δt , with the same inter-observation interval T , will lead to an algorithm which is prohibitively costly. Therefore, it can be argued that such proposals are consistent for any step-size, Δt , that may be used in practice.

D Raw Results

In this appendix we include, for completeness, the raw relative effective sample sizes (as defined by (4.2)) and the average execution times for each proposal and each combination of (T, y_1^T) for the Lotka-Volterra and gene-expression diffusions detailed in section 4, and for the birth-death diffusion detailed in appendix B. Recall that, for each combination of (T, y_1^T) , we simulated 1,000,000 independent skeleton paths using five different proposals; the MDB of Durham and Gallant (2002), the residual-bridge proposal of Whitaker et al. (2016) with the two choices for ξ_t , RB^{ODE} and RB^{LNA} , and the residual-bridge proposal introduced in this paper with the same two choices for ξ_t , $\overline{\text{RB}}^{\text{ODE}}$ and $\overline{\text{RB}}^{\text{LNA}}$. For each proposal and each combination of (T, y_1^T) we calculated the normalised weights for each of the 1,000,000 paths according to (2.1) and used these to calculate the relative effective sample size (Rel. ESS) defined by (4.2). We also noted the average execution time (wall time) in seconds over ten identical runs for each algorithm. The relative effective sample sizes and average execution times can be seen, respectively, in figures 10 and 12 for the birth-death diffusion, and in figures 11 and 13 for the Lotka-Volterra and gene-expression diffusions.

We note that whilst a small relative effective sample size is not ideal, indicating as it does a relatively poor proposal, if this proposal is the best among its competitors then it is still the best option, and, with a large enough absolute effective sample size, inference which utilises this proposal can still be performed accurately.

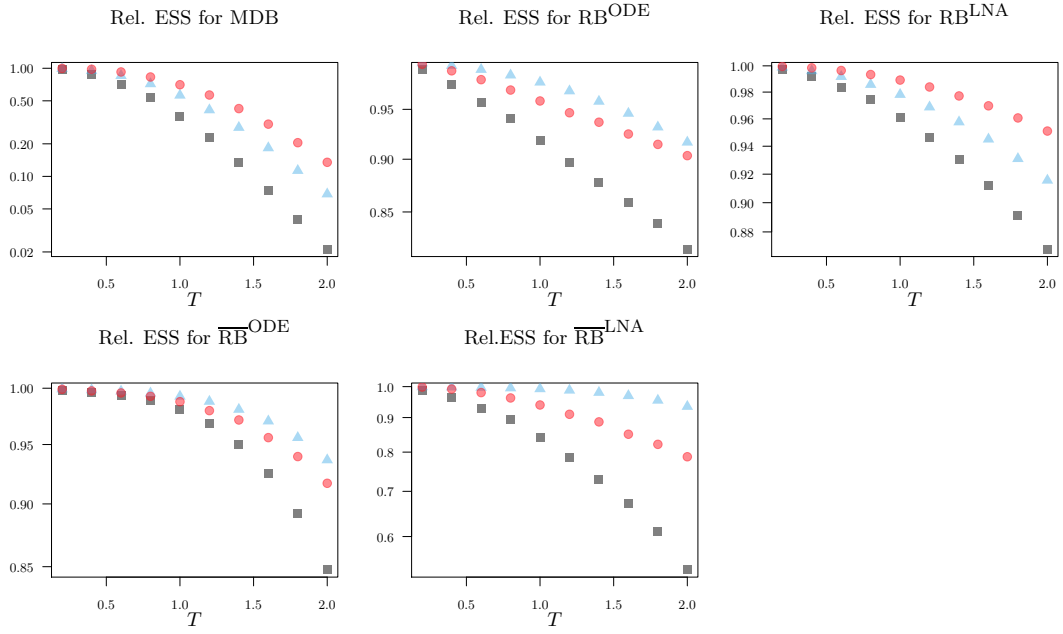


Figure 10: Plots of the relative effective sample sizes (as defined by (4.2)) for five proposals and for a variety of combinations of (T, y_1^T) corresponding to the birth-death diffusion. The observations, y_1^T , corresponding to the 50% quantile of the simulated observations are denoted with blue triangles, the observations corresponding to the 5% quantile are denoted with grey boxes, and the observations corresponding to the 95% quantile are denoted with red circles.

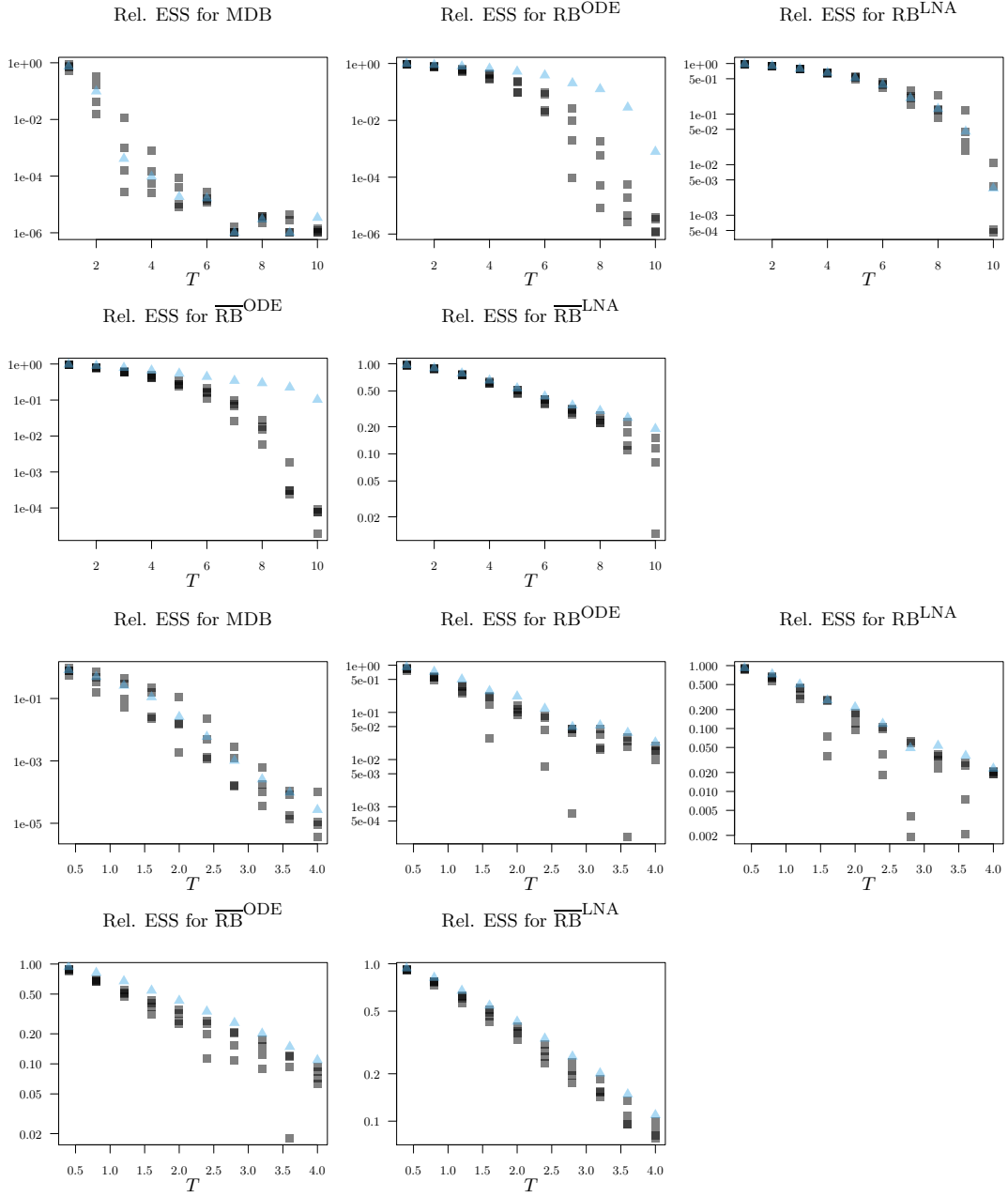


Figure 11: Two sets of five plots of the relative effective sample sizes (as defined by (4.2)) where the top (respectively bottom) five plots are the effective sample sizes for five proposals and for a variety of combinations of (T, y_1^T) corresponding to the Lotka-Volterra (respectively gene-expression) diffusion. The observations, y_1^T , corresponding to the mean of the simulated observations are denoted with blue triangles whereas the observations corresponding to the four 90% quantiles along the axes of the principal components are denoted with grey boxes.

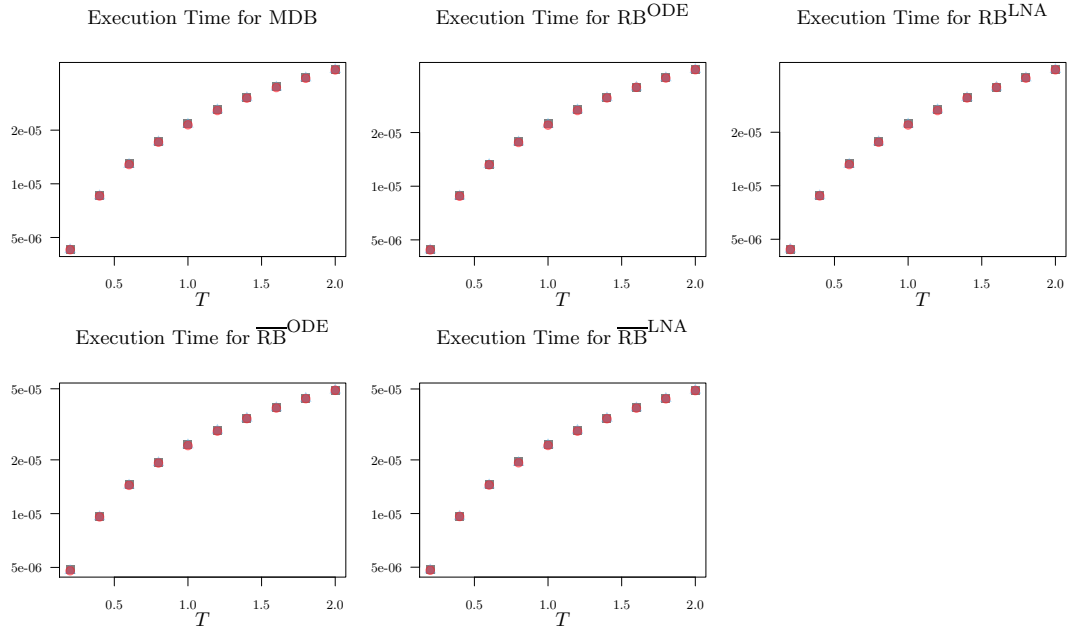


Figure 12: Plots of the average execution times for five proposals and for a variety of combinations of (T, y_1^T) corresponding to the birth-death diffusion. The observations, y_1^T , corresponding to the 50% quantile of the simulated observations are denoted with blue triangles, the observations corresponding to the 5% quantile are denoted with grey boxes, and the observations corresponding to the 95% quantile are denoted with red circles.

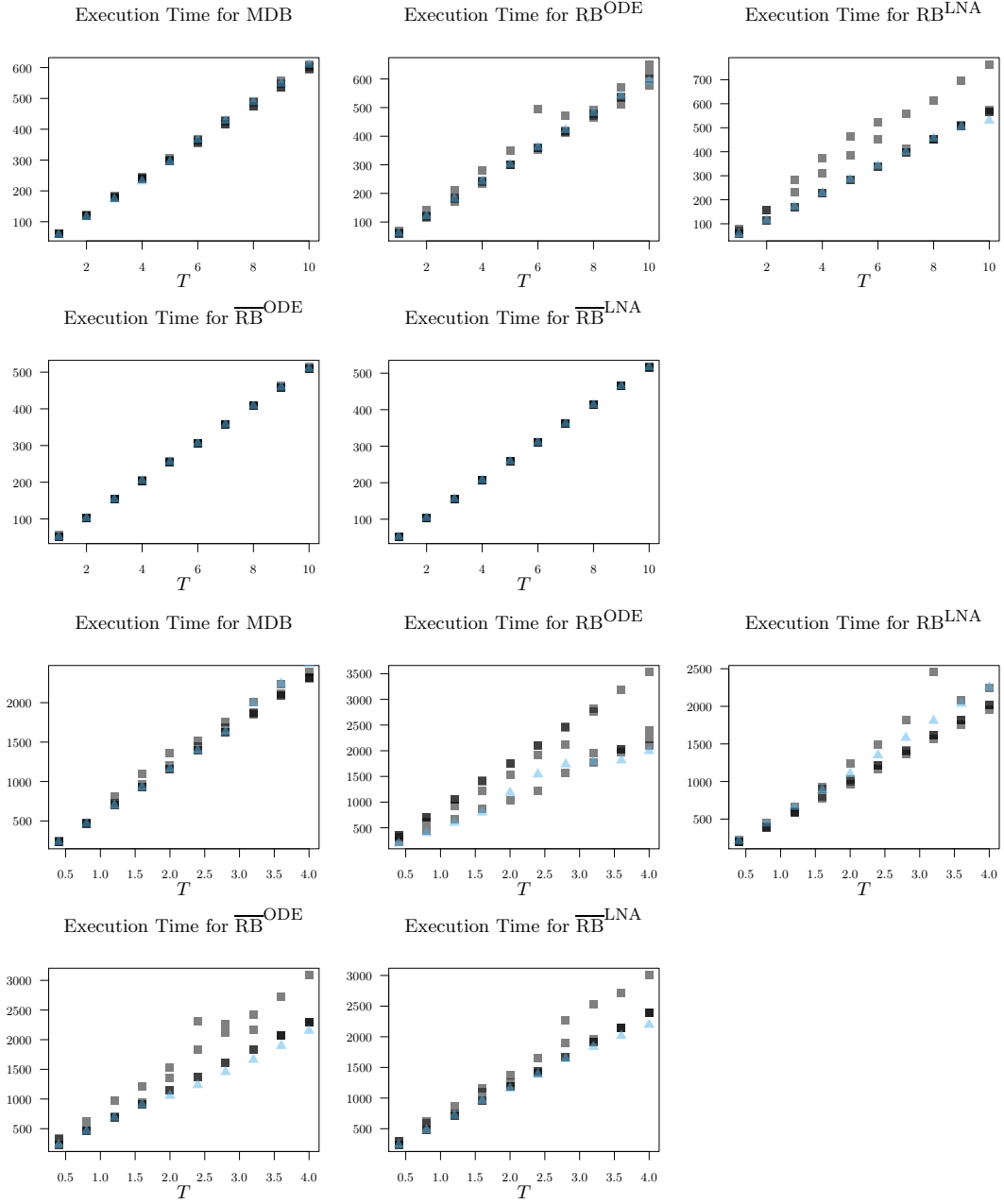


Figure 13: Two sets of five plots of the average execution times where the top (respectively bottom) five plots are the effective sample sizes for five proposals and for a variety of combinations of (T, y_1^T) corresponding to the Lotka-Volterra (respectively gene-expression) diffusion. The observations, y_1^T , corresponding to the mean of the simulated observations are denoted with blue triangles whereas the observations corresponding to the four 90% quantiles along the axes of the principal components are denoted with grey boxes.